Zur Frage angemessener Methodenstrategien bei der Vorhersage des Studienerfolgs

Eine vergleichende Evaluation von linearen und nichtlinearen Modellen der Datenanalyse

Hartmut Schlattmann

1970

ZUR FRAGE ANGEMESSENER METHODENSTRATEGIEN BEI DER VORHERSAGE DES STUDIENERFOLGS

Eine vergleichende Evaluation von linearen und nichtlinearen Modellen der Datenanalyse

D I S S E R T A T I O N

zur Erlangung des Grades eines

Doktors der Philosophie
der Philosophischen Fakultät
der Universität des Saarlandes

vorgelegt von
Hartmut Schlattmann
aus Bielefeld

Dekan:	Prof. Dr. L. Kötter
Berichterstatter:	Prof. Dr. P. Orlik Prof. Dr. W. Tack
Tag der letzten Prüfung:	3. März 1978

UNIV.
LEIDEN
BIBL.

INHALTSVERZEICHNIS

		Seite
1.	Vorwort	3
2.	Einleitung	5
2.1	Ausgangspunkt	5
2.2	Fragestellung	6
2.3	Datenerhebung	11
2.3.1	Die befragten Personen	11
2.3.2	Die erfaßten Merkmale	14
2.4	Reichweite der Aussagen	17
3.	Multiple Regressionsanalyse	19
3.1	Eigenschaften des Verfahrens und notwendige Voraussetzungen der Daten	20
3.2	Die Anwendbarkeit des Verfahrens auf die vorliegenden Daten	23
3.2.1	Skalenqualität	23
3.2.2	Modellvoraussetzungen	27
3.3	Ergebnisse	30
3.4	Diskussion	33
3.4.1	Abiturnoten	33
3.4.2	Die "harten" Prädiktoren	46
3.4.3	"Alle" Prädiktoren	49
3.4.4	Schlußfolgerungen	50
4.	Weitere Untersuchungen im Rahmen des linearen Modells	52
4.1	Prüfung des Kriteriums "Zensur im Zwischenexamen" auf Gleichabständigkeit der Skalenstufen und auf Dimensionalität	52
4.1.1	Rationale	52
4.1.2	Ergebnisse und Diskussion	55
4.2	Unterscheidung der Studienfächer aufgrund der Abiturnoten	59
4.2.1	Ergebnisse	60
4.2.2	Diskussion	67

		Seite
4.3	Gleichzeitige Vorhersage von Studienfach und Studienerfolg	71
4.3.1	Ergebnisse	72
4.3.2	Diskussion	72
4.4	Vorschlag eines Modells, das gleichzeitig Studienerfolg und Studienabbruch als Kriterien einbezieht	80
4.5	Zusammenfassung	82
5.	Erweiterungen des linearen Modells	83
5.1	Moderatoranalyse	84
5.2	Moderatoranalyse nach Saunders	91
5.3	Analyse der Residuen	92
5.4	Moderatoranalyse nach Ghiselli	97
5.5	Diskussion und Zusammenfassung	100
6.	Die Kontrastgruppenanalyse (AID)	102
6.1	Komplexe Merkmalzusammenhänge	102
6.2	Das Verfahren der Kontrastgruppenanalyse (AID)	105
6.3	Ergebnisse	109
6.4	Diskussion	117
6.5	Kreuzvalidierung der Ergebnisse von multiplen Regressionsanalysen und von Kontrastgruppenanalysen	120
6.6	Die Analyse von Residuen mit der Kontrastgruppenanalyse	122
7.	Die Konfigurationsfrequenzanalyse (KFA)	123
7.1	Konfigurationsfrequenz - und Interaktionsstrukturanalyse	125
7.2	Ergebnisse	127
7.3	Diskussion	130
7.4	Die Analyse von Residuen mit der KFA	132
8.	Schlußfolgerungen	134
	Literaturverzeichnis	139

1. Vorwort

In dieser Arbeit soll versucht werden, den Stand des Wissens über die Möglichkeiten einer Vorhersage des Studienerfolgs - aufgefaßt als Zensur im Zwischenexamen - zu erweitern. Die Hochschulforschung innerhalb dieses Bereichs weist bereits eine ganze Reihe von Untersuchungen auf, die zu einigermaßen übereinstimmenden inhaltlichen Befunden kommen. Jedoch bleibt eine größere Anzahl methodologischer Probleme, die teils in den einschlägigen Arbeiten angesprochen werden, teils ungenannt bleiben oder sich erst in der zusammenfassenden Bewertung ergeben, umstritten bzw. offen. Von besonderem Interesse erscheint dabei vor allem die Vermutung, daß die unzureichende Vorhersagbarkeit des Studienerfolgs, die von praktisch allen Untersuchern konstatiert wird, durch die Beschränkung in den Techniken der Datenanalyse auf lineare, kompensatorische Modelle zustandekommt. Möglicherweise könnte die Vorhersage des Studienerfolgs wesentlich besser sein, wenn man solche Methoden benutzt, die nicht nur lineare, sondern auch partiell nichtlineare und völlig adimensionale, konfigurationale Effekte berücksichtigen. Hauptzweck dieser Studie ist deshalb der kritische Vergleich verschiedener linearer, partiell nichtlinearer und völlig adimensionaler, konfigurationaler Methoden der Datenanalyse in exemplarischer Anwendung auf einen vorliegenden Satz psychologischer Daten. Die Eignung der verschiedenen Analysetechniken zur Transformation der in diesem Datensatz enthaltenen Strukturen in eine für den Untersucher unmittelbar verständliche Form soll einer vergleichenden Evaluation unterzogen werden.

Zusätzlich sollen weitere methodische Probleme - z.B. ob sich parametrische Methoden grundsätzlich auf Daten mit der üblichen zwischen Ordinal- und Intervallniveau liegenden Meßqualität anwenden lassen und ob Studierende der verschiedenen Studienfächer in einer gemeinsamen Analyse gleichzeitig untersucht werden können - innerhalb dieser

exemplarischen Demonstration der Einsatzmöglichkeiten verschiedener Methoden der Datenanalyse einer Klärung nähergebracht werden. Insgesamt soll mit der angestrebten Auswahl und Bewertung der empfehlenswerten Methoden und ihren Einsatzmöglichkeiten ein Beitrag zur Forschungsmethodologie des Problembereichs "Vorhersage des Studienerfolgs" geleistet werden.

Die in der vorliegenden Studie als Beispiel analysierten Daten entstammen einem von der Universität Saarbrücken durchgeführten größeren Forschungsprojekt zu "Orientierungsprobleme und Erfolgsbeeinträchtigung bei Studierenden". Eine Beschreibung der Ziele und der Datenerhebung dieses Projekts, sowie eine Präsentation der Befunde findet man in APENBURG, GROSSKOPF und SCHLATTMANN (1974). Von den darin genannten Mitarbeitern der Forschungsgruppe, auf deren Bemühungen auch die vorliegende Studie basiert, seien stellvertretend die Leiter Dr. H. Saterdag und Dipl. Psych. E. Apenburg und als Mitglieder des wissenschaftlichen Beirates die Professoren Boesch, Fisch, Kötter, Orlik und Strauch genannt.

Diese Arbeit wurde von Prof. Dr. Orlik betreut, dem ich für wichtige Hinweise vor allem in bezug auf die grundsätzliche Anlage zu besonderem Dank verpflichtet bin.

Die dieser Studie zugrunde liegenden Berechnungen hätten nicht ohne die Möglichkeiten eines leistungsfähigen Rechenzentrums wie dem der Universität Köln und ohne die bereitwillige Hilfe seiner Mitarbeiter - stellvertretend zu nennen Dipl. Math. D. Kögel- geleistet werden können.

2. Einleitung

2.1 Ausgangspunkt

Seit einiger Zeit erfahren Fragen der Hochschulzugangsregelung und in Zusammenhang damit Fragen der Vorhersagbarkeit des Studienerfolgs wohl im Zusammenhang mit gesellschaftlichen Veränderungen und politischen Diskussionen wachsende Aufmerksamkeit. Dadurch angeregt stieg auch die Menge entsprechender empirischer Untersuchungen, wenn sich auch deren Anzahl - selbst unter Berücksichtigung der unterschiedlichen Bevölkerungszahlen und des unterschiedlichen Umfangs der Bildungseinrichtungen - nicht im entferntesten mit der Anzahl entsprechender Untersuchungen in den USA vergleichen läßt. So sind die Befunde der deutschen Untersuchungen durchaus noch zu überschauen (in den USA ist man dagegen meist gezwungen, auf die periodisch erscheinenden Übersichten zurückzugreifen); sie wurden in jüngster Zeit von TROST (1975) zusammenfassend dargestellt. Für diejenigen deutschen Untersuchungen, die sich der Examenszensuren als Kriterium des Studienerfolgs bedienen - vor allem diese sind ja für die vorliegende Studie relevant - lassen sich m.E. vor allem folgende drei Feststellungen treffen:

1. Die Vorhersagevaliditäten sind schwach; sie liegen zu niedrig für eine Vorhersage im Einzelfall und zu niedrig zur Begründung von Selektionsmaßnahmen. Sie sind deutlich geringer als die in den USA gefundenen, wo z.B. die multiplen Korrelationen zwischen Noten der high school und der ersten oder zweiten Studienjahre zu 65 % zwischen 0,45 und 0,65, im Mittel bei 0,55 lagen (TROST, 1975, aufgrund von 310 Validitätsstudien). Zur Erklärung der Diskrepanz zwischen deutschen und amerikanischen Ergebnissen werden von TROST mehrere Unterschiede des amerikanischen und deutschen Schulsystems angeführt.

2. In den deutschen Untersuchungen werden vorwiegend nur Prädiktoren eines engen Bereichs verwendet; meist werden mit den Abiturnoten und wenigen, sich anbietenden, zusätzlichen Merkmalen wie z.B. Geschlecht oder Schultyp

vor allem diejenigen Merkmale untersucht, die in der Praxis der Zugangsregelung Verwendung finden. Es ist jedoch plausibel und wird durch einzelne Untersuchungen nahegelegt, daß die Vorhersage durch weitere Prädiktoren verbessert werden könnte.

Auch die Erfassung des Kriteriums muß insofern als unvollständig gewertet werden, als praktisch durchgängig jene Studierenden nicht in die Untersuchungen einbezogen wurden, die das Studium überhaupt oder das Studium des betreffenden Faches abgebrochen hatten, ohne sich einer Prüfung zu unterziehen. Zur Vermeidung dieses Nachteils wären Längsschnittuntersuchungen notwendig; einen Eindruck von den dabei auftauchenden Schwierigkeiten geben AMELANG und HOPPENSACK (1977 a und 1977 b).

3. Es wurden in Form der einfachen oder multiplen Korrelationsrechnung in allen Untersuchungen gleichartige Methoden der Datenanalyse bzw. Inferenzstatistik angewendet. Dabei wurde insgesamt gesehen weder geprüft, ob die Voraussetzungen für die Anwendbarkeit dieser Verfahren gegeben sind, noch ob vielleicht mit andersartigen Verfahren andere, bessere Möglichkeiten der Vorhersage zu erzielen sind.

2.2 Fragestellung

Die vorliegende Studie will verschiedene Techniken der Datenanalyse einer vergleichenden Evaluation unterziehen. Grundsätzlich können Analysetechniken auf verschiedenen Wegen erprobt und verglichen werden: Man kann sie auf konstruierte Datenbeispiele mit extrem einfacher Struktur, auf nach komplexeren Zufallsmodellen generierte Daten oder auf konkrete, z.B. zu Forschungszwecken erhobene Daten, bei denen es auf die tatsächliche Produktion von Hypothesen und Entscheidungen ankommt, anwenden, um dabei die Eignung der Verfahren zur Transformation der Datenstrukturen in für den Untersucher verständliche und die Sachverhalte erhellende Form zu prüfen und zu vergleichen.

Bei den ersten beiden der aufgezählten Möglichkeiten bieten die Konstruktionsgesichtspunkte der Daten quasi ein Kriterium, zu dem das zu bewertende Verfahren in Beziehung gesetzt wird. Das Verfahren ist "valide", wenn es die in die Konstruktion der Daten hineingegebenen Prinzipien in angemessener Weise aufzudecken imstande ist. Man kann dabei sehr viele unterschiedlich strukturierte Datensätze generieren und jeweils die Transinformation durch die betreffende Analysetechnik untersuchen.

Unter der letzten der aufgezählten Möglichkeiten - bei der Bewertung der Analysemethoden in beispielhafter Anwendung auf einen bestimmten Datensatz, so wie in dieser Studie geplant - liegt der Validierungsgesichtspunkt zum einen im Vergleich der verschiedenen Techniken untereinander. Dabei spielt z.B. in dieser Arbeit die multiple Regressionsanalyse als das übliche und in vielen Anwendungen erprobte Verfahren die Rolle eines Bezugsrahmens, an dem die Leistungen der anderen, bisher nicht in diesem Bereich erprobten Verfahren gemessen werden. Zum anderen kann die "Validität" der Verfahren daran gemessen werden, inwieweit sie überhaupt in den Daten enthaltene - a priori unbekannte - Strukturen aufzudecken vermögen. Dabei kann es kein Kriterium in Form einer klar definierten Datenstruktur geben, sondern es muß die Entdeckung jeglicher Effekte, die sich innerhalb des verwendeten Verfahrens sichern lassen, zunächst per se als positiv und dann entsprechend dem dadurch erzielten Nutzen für die Klärung der Probleme bewertet werden. Die schließlich vorgenommene Evaluation der Analysetechniken kann nur für die zugrundegelegten Daten, sowie für Daten mit gleichartiger innerer Struktur als gültig betrachtet werden. Zur Bewertung der Stichprobenqualität der hier verwendeten Daten vgl. 2.4.

Die Wahrscheinlichkeit dafür, daß die in dieser Arbeit vorgefundenen Datenstrukturen für den Bereich der Vorhersage des Studienerfolgs repräsentativ sind, wird durch die getrennte Analyse der einzelnen Studienfächer in positiver Weise beeinflußt. Man kann nämlich die vorliegende Stichprobe von Studierenden verschiedener Fächer auch als Gruppe von Stichproben aus jeweils einzelnen

Studienfächern auffassen. Danach können die Bewertungen der Analyseverfahren, weil sie in verschiedenen Stichproben zunächst unabhängig voneinander vorgenommen wurden, als besser abgesichert betrachtet werden, als wenn sie auf lediglich einer einzelnen Stichprobe basierten.

Für die Durchführung der beabsichtigten Analysen, für die Generalisierbarkeit der Befunde und für die wünschenswerte Kreuzvalidierung muß man als Voraussetzung u.a. eine recht umfangreiche Stichprobe fordern. Einzelheiten über die verwendeten Daten enthält 2.3. Hier sei nur festgestellt, daß im Vergleich zu den erfaßten ca. 2400 Studierenden deutsche Untersuchungen sehr selten vergleichbare und lediglich in einem mir bekannten Fall (SCHÜTZ, 1968) eine umfangreichere Stichprobe aufwiesen. Das Spektrum der Prädiktoren umfaßte nur in wenigen Fällen deutlich mehr Merkmale als die Abiturnoten allein. Auch amerikanische Untersuchungen verwendeten kaum jemals größere Stichproben bei vergleichbarer Spannweite der Prädiktoren, was angesichts des Aufwands bei der Erhebung großer Datenmengen (wie z.B. der in 2.3 beschriebenen Daten) verständlich erscheint. So dürfte die Hoffnung auf bessere und umfangreichere Stichproben - auch wenn diese durchaus wünschenswert und eigentlich notwendig wäre - wenig realistisch sein.

Die Bezeichnung der methodischen Verfahren als Verfahren der Datenanalyse folgt der von TUKEY (1962) in seinem Artikel "the future of data analysis" aufgewiesenen Richtung. So ist diese Untersuchung stärker an der Beantwortung von Fragen aus dem behandelten inhaltlichen Gebiet als an der Mathematik der verwendeten Methoden interessiert, die Verfahren werden eher als "indication procedures" statt als "conclusion procedures" aufgefaßt und es wird im Interesse des Bemühens um richtige Antworten in Kauf genommen, sich gelegentlich auch einmal zu irren.(vgl. auch: COOLEY und LOHNES, 1971, MEREDITH et al., 1974). Bei den verwendeten Verfahren stehen solche mit einem gut ausgearbeiteten statistischen Konzept wie z.B. der multiplen Regressionsanalyse neben einem Verfahren wie der Kontrastgruppenanalyse, für

das statistische Konzepte höchstens zu einem sehr geringen Ausmaß bestehen. Wenn statistische Tests durchgeführt werden - und dies geschieht überall dort, wo es möglich ist und nützlich erscheint -, sollen sie mehr in beschreibendem Sinn im Hinblick auf praktische Bedeutsamkeit als im Sinn statistischer Entscheidungen interpretiert werden (vgl. z.B. SIGNORELLI, 1974).

Im einzelnen wird sich der Inhalt dieser Studie um folgende fünf Gesichtspunkte zentrieren, die die zu evaluierenden Methoden der Datenanalyse in einer Reihe anordnen, die vom linearen Modell (Regressions- und Diskriminanzanalyse), über Abwandlungen dieses Modells (durch Variablentransformationen und Moderatoransätze), zu neuartigen Modellen mit der Berücksichtigung nichtlinearer und konfigurationaler Effekte (Kontrastgruppen- und Konfigurationsfrequenzanalyse) reicht. Die Inhalte der fünf Gesichtspunkte werden durch in Stichworten angedeutete Einzelfragen illustriert; jeder Einzelfrage ist in Klammern der entsprechende Teil der Arbeit zugeordnet, so daß diese Aufzählung auch als Ergänzung des Inhaltsverzeichnisses zur klareren Darstellung der inneren Abhängigkeiten dienen kann.

1. Einsatz der multiplen Regressionsanalyse als Standardverfahren und als Bezug für die Bewertung der anderen Verfahren; die Punkte a - c beziehen sich auf das rein lineare Modell, d - e schließen die Möglichkeit nichtlinearer Effekte durch Variablentransformationen ein

   a. Prüfung der grundsätzlichen Anwendbarkeit des Modells bezüglich der Meßqualität der Daten und Prüfung der konkreten Anwendbarkeit bezüglich der Modellannahmen (3.1, 3.2, 5.3)

   b. Untersuchung, ob die Studienfächer gemeinsam analysierbar sind (3.5 bis 3.5, 5.3)

c. Darstellung der inhaltlichen Ergebnisse des gesamten Prädiktorenspektrums (3.3 bis 3.5, 5.3)

d. Suche nach kurvilinearen Effekten einzelner Prädiktoren (3.2.2 und 5.3)

e. Untersuchung von Wechselwirkungen analog der Varianzanalyse (5.2)

f. Prüfung, ob die Auffassung des Kriteriums in der Regressionsanalyse als eindimensionale Intervallskala eine optimale Anpassung des Modells an die Daten ermöglicht (4.1)

2. Erweiterung des grundlegenden inhaltlichen Ansatzes - Beziehung der Validität der Abiturnoten auf weitere Merkmale als die Examenszensuren allein - mit Hilfe des linearen Modells

a. Unterscheidung der Studienfächer aufgrund der Abiturnoten (4.2)

b. Vorhersage gleichzeitig des Examenserfolgs und des Studienfachs, Unterscheidung von zwei Komponenten der Validität der Abiturnoten (4.3)

c. Vorschlag eines Ansatzes zur gleichzeitigen Vorhersage des Examenserfolgs und Studienabbruchs (4.4)

3. Erweiterung des linearen Modells durch Techniken der Moderatoranalyse

a. Überlegungen zur Situation bei der Entwicklung und Anwendung von Moderatortechniken (5.1)

b. Verbesserung der Vorhersage durch die multiple Regressionsanalyse mit Hilfe der Moderatoranalyse nach SAUNDERS (5.2)

c. Verbesserung der Vorhersage durch die multiple Regressionsanalyse mit Hilfe der Moderatoranalyse nach GHISELLI (5.3)

4. Ersatz des linearen Modells durch die Kontrastgruppenanalyse (AID); Berücksichtigung linearer, dimensional nichtlinearer und unter bestimmten Umständen völlig adimensionaler, konfigurationaler Effekte; zu erwartende relativ gute Effizienz durch Berücksichtigung quantitativer Skaleneigenschaften bei

den analysierten Merkmalen

a. Effekte der Kontrastgruppenanalyse in Vergleich zu Effekten der Regressionsanalyse, Kreuzvalidierung von Ergebnissen beider Verfahren (6.1 bis 6.5)

b. Kombination der Regressions- und Kontrastgruppenanalyse; Prüfung, ob die Kontrastgruppenanalyse neuartige Effekte entdeckt, die bei der Regressionsanalyse verborgen blieben (6.6)

5. Ersatz des linearen Modells durch die Konfigurationsfrequenzanalyse (KFA) bzw. Interaktionsstrukturanalyse; Berücksichtigung jeglicher linearer und nichtlinearer, konfigurationaler Effekte; zu erwartende relativ geringe Effizienz durch Zurückführung der Skalenqualität aller beteiligten Merkmale auf Nominalniveau

a. Effekte der KFA im Vergleich zur Regressionsanalyse und Kontrastgruppenanalyse (7.1 bis 7.3)

b. Kombination von Regressionsanalyse und KFA; Prüfung, ob die KFA neuartige Effekte entdeckt, die bei der Regressionsanalyse verborgen blieben (7.4)

2.3 Datenerhebung

2.3.1 Die befragten Personen

In der vorliegenden Arbeit wurden, wie bereits im Vorwort erwähnt, Daten aus einer Untersuchung verwendet, die im Rahmen eines größeren Forschungsprojektes der Universität Saarbrücken zu "Orientierungsproblemen und Erfolgsbeeinträchtigung bei Studierenden" durchgeführt wurde.

Es handelte sich dabei um eine umfangreiche Befragung von Studierenden der Universität Saarbrücken. Der Fragebogen enthielt 142 Items, die sich auf Elternhaus, Verhalten und Leistungen in der Schule, Bewertung der Studienbedingungen und des Lehrbetriebs und auf studienbezogene Verhaltensweisen und individuelle Merkmale der Studierenden bezogen.

Der Fragebogen wurde der Gesamtheit der im WS 1972/73 an der Universität des Saarlandes immatrikulierten deutschen Studenten zugeschickt. Von den 9526 angeschriebenen Studenten schickten 4440 den Fragebogen - teilweise nur unvollständig ausgefüllt - zurück. Dies entspricht einer Rücklaufquote von 46,6 %. 3772 der zurückerhaltenen Fragebögen waren auswertbar, damit repräsentierte die erfaßte Teilgesamtheit bzw. Stichprobe 39,6 % der Population.

Beim Vergleich von Stichprobe und Grundgesamtheit zeigte sich, daß männliche und weibliche Studierende erwartungsgemäß repräsentiert waren, daß die einzelnen Studienfächer unterschiedliche Rücklaufquoten aufwiesen - zwischen 25,6 % und 47,1 % - und daß außerdem die Studenten mit niedriger Semesterzahl in unserer Stichprobe überrepräsentiert waren.

Die Daten der Erhebung wurden in einem 1974 erschienenen Bericht vollständig präsentiert (vgl. APENBURG, GROSSKOPF & SCHLATTMANN, 1974).

In die vorliegende Untersuchung wurden alle jene Studierende einbezogen, die bereits ein Zwischenexamen abgelegt hatten, so daß die darin erzielten Noten als Kriterium des Studienerfolgs dienen konnten. Zu den 1661 Studenten, für die dieses zutraf, kamen als weitere Gruppe 354 Studierende, die zum Zeitpunkt der Haupterhebung zwar noch kein Zwischenexamen abgelegt hatten, von denen aber die entsprechenden Zensuren in einer Nacherhebung im SS 1976 erfaßt werden konnten. Schließlich war in dieser Nacherhebung noch eine Gruppe von 548 Studierenden enthalten, die zwar ein Zwischenexamen abgelegt hatten, jedoch in der Haupterhebung überhaupt nicht enthalten waren; diese Gruppe von Studierenden weist deshalb einen eingeschränkten Variablensatz - im wesentlichen die Abiturnoten und das Kriterium - auf (vgl. 2.3.2). So liegen also von 1661 + 354 = 2015 Studierenden - bis auf einzelne fehlende Angaben - vollständige, und von weiteren 348 Studierenden unvollständige Datensätze vor, d.h. die gesamte untersuchte Gruppe umfaßt ca. 2360 Studierende. Über Prüfungen auf Gleichwertigkeit der zu den unterschiedlichen Zeitpunkten erhobenen Daten und über die Behandlung der fehlenden Angaben wird später zu reden sein.

Bei den Angaben der Fachrichtungen wurden jeweils Betriebswirtschaft und Volkswirtschaft, sowie Mathematik und Informatik zusammengefaßt, so daß die sich ergebenden Fachgruppen folgendermaßen repräsentiert waren:

Tabelle 2.1: Die zahlenmäßige Repräsentation der einzelnen Studienfächer

Fach	Abkürzung	Umfang
Betriebswirtschaft/ Volkswirtschaft	BWL / VWL	679
Medizin	Med	425
Mathematik / Informatik	Math / Inf	264
Deutsch	Deutsch	196
Französisch	Franz	167
Chemie	Chemie	97
Wirtschaftspädagogik	Wi - Päd	102
Physik	Physik	106
Restliche Fächer		327
Summe		2363

2.3.2 Die erfaßten Merkmale

Von den ursprünglich in der Untersuchung enthaltenen Merkmalen wurde eine Auswahl nach den folgenden zwei Gesichtspunkten getroffen:

1. Mit welchen Variablen kann der Studienerfolg vorhergesagt werden?
2. Welche Variablen können zur Erhöhung der Zusammenhänge des Kriteriums mit den nach 1. ausgewählten Merkmalen durch Suppressor- oder Moderatoreffekte beitragen?

Es finden lediglich solche Merkmale als Prädiktoren Verwendung, die vermutlich bereits vor Aufnahme des Studiums ausgeprägt waren, die sich zumindest nicht erst während des Studiums entwickelten (wie z.B. Arbeitshaltungen oder Einstellungen zum Studieren). Als Problem tritt dabei allerdings bei einer vorwiegend im Querschnitt durchgeführten Untersuchung wie der vorliegenden auf, daß bestimmte Prädiktoren quasi rückwirkend vom Kriterium beeinflußt werden können, d.h. daß sich Studenten ausgehend von guten Examenszensuren oder Erfolg in bestimmten Veranstaltungen auch in bestimmten Persönlichkeitsmerkmalen - es waren ja fast alle Merkmale in der Selbstbeurteilung der Studierenden erhoben worden - günstigere Ausprägungen zuschreiben als es bei Mißerfolgen der Fall wäre. Dieser Effekt ist weniger bei klar definierten und auf eindeutigen Fakten beruhenden oder subjektiv wenig mit dem Erfolg im Studium zusammenhängenden Merkmalen wie z.B. Abiturnoten oder dem persönlichen Verhältnis zu den Lehrern und entsprechend stärker bei den diffuser definierten und subjektiv mehr mit dem Studienerfolg zusammenhängenden Merkmalen wie z.B. der allgemeinen intellektuellen Leistungsfähigkeit zu erwarten. Entsprechend dieser Bewertung wird bei den nachfolgenden Aufzählungen zwischen "harten" und "weichen" Prädiktoren unterschieden.

Folgende 14 Merkmale machen die "harten" Prädiktoren aus, die jeweils mit ihren Kurzbezeichnungen und Erklärungen aufgeführt sind:

Kurzbezeichnung	Erklärung
Deutsch	
Englisch bzw. 1.Fremdsprache	
Erdkunde	
Mathematik	
Physik	
Biologie	Abiturzensuren in den entsprechenden Fächern
Musik	
Geschlecht	
Schicht	bildungsrelevanter Sozialstatus, berechnet aus dem Schulabschluß beider Eltern und dem Beruf des Vaters bzw. der Mutter (nach BAUER, 1972)
Vertrauenspersonen[1)]	mit wem (Vater, Mutter, Geschwister, Lehrer oder niemand) hat man sich während der Schulzeit über persönliche Probleme unterhalten
Depression[1)]	Faktorvariable aus einer Auswahl von MMPI-Fragen
Verhältnis zu den Lehrern[1)]	quantitative Beurteilung des Verhältnisses zu den Lehrern
Kontaktfindung[1)]	quantitative Angabe, wie leicht es fiel, zu Beginn des Studiums Kontakt zu Kommilitonen zu bekommen (dieses Merkmal wurde teilweise den harten Prädiktoren als Indikator für Extraversion angefügt, obwohl es aus dem Bereich des Studiums stammt)

1) Dieses Merkmal ist bei den 348 Studierenden, die nur in der Nacherhebung (vgl. 2.3.1) erfaßt worden waren, nicht vorhanden.

Fachinteresse[1)]	Übereinstimmung des Studienfachs mit den interessant gefundenen Schulfächern (ja oder nein): so definiert für die Studienfächer Math/Inf, Deutsch, Französisch, Chemie, Physik; bei Medizin wenn Interesse an Biologie oder Chemie

Die 11 "weichen" Prädiktoren sind folgende:

Kurzbezeichnung	Erklärung
Intelligenz	Selbsteinschätzung der intellektuellen Leistungsfähigkeit
Mathem. Begab.[1)]	Selbsteinschätzung der mathematischen Begabung
Sprachl. Begab.[1)]	Selbsteinschätzung der sprachlichen Begabung
Naturw. Begab.[1)]	Selbsteinschätzung der naturwissenschaftlichen Begabung
Künstl. Begab.[1)]	Selbsteinschätzung der künstlerischen Begabung
Technische Begab.[1)]	Selbsteinschätzung der technischen Begabung
Kaufm. Begab.[1)]	Selbsteinschätzung der kaufmännischen Begabung
Prakt. Begab.[1)]	Selbsteinschätzung der praktischen Begabung
	Faktorvariablen der Leistungsmotivation in der Selbstbeurteilung: (nach FISCH et al., 1970)
LM - HE	Hoffnung auf Erfolg
LM - FE	Furcht vor Mißerfolg
LM - N	Neutral

1) Dieses Merkmal ist bei den 348 Studierenden, die nur in der Nacherhebung (vgl. 2.3.1) erfaßt worden waren, nicht vorhanden.

Zusätzlich wurden noch die Merkmale Fachsemesterzahl und Prüfungsangst[1] (Selbstbeurteilung der Angst vor Prüfungen und Klausuren) wegen der Wahrscheinlichkeit nichtlinearer und nichtkompensatorischer Effekte in die Untersuchung einbezogen.

2.4 Reichweite der Aussagen

Es ist hier zunächst vor allem von der verwendeten Stichprobe zu sprechen, von der Beschränkung auf eine Universität (Saarbrücken) und von der Voraussetzung der freiwilligen Kooperation der Studierenden. Im Vergleich zu anderen deutschen und auch amerikanischen Untersuchungen kann man feststellen, daß diese nach den Regeln der Stichprobenerhebung ungünstige Situation häufig besteht. Die Beschränkung auf eine oder wenige Hochschulen kann als Norm gelten, und wenn die Mitarbeit der Studenten nicht auf freiwilliger Basis erfolgte, handelte es sich meist um Erhebungen - dann vorwiegend Vollerhebungen - aus Unterlagen der Prüfungsämter oder anderer Institutionen (in den USA meist Zulassungsinstitutionen der Hochschulen). Eine der wenigen Ausnahmen ist z.B. die Untersuchung von BARTENWERFER und GIESEN (1973) an einer Zufallsstichprobe von Gymnasiasten. Man muß also die inhaltlichen Aussagen dieser Untersuchung als beispielhaft für eine auf freiwilliger Mitarbeit basierende Auswahl von Studierenden der Universität Saarbrücken, die im Zeitraum von 1972 - 1975 eingeschrieben waren und ein Zwischenexamen abgelegt hatten, bewerten.

Durch eine entsprechende oder ähnliche Situation wird in vielen Untersuchungen die Möglichkeit einer Kreuzvalidierung mit Daten anderer Untersucher verhindert. Oder eine durchgeführte Kreuzvalidierung nimmt eine andere Bedeutung als die übliche an, weil es sich nicht um Stichproben aus derselben Grundgesamtheit handelt. Zum Teil in diesem Sinn

---

1) Dieses Merkmal ist bei den 348 Studierenden, die nur in der Nacherhebung (vgl. 2.3.1) erfaßt waren, nicht vorhanden.

ist wohl auch die Feststellung verschiedener Autoren (z.B. COOLEY und LOHNES, 1971) zu sehen, daß multiple Korrelationen bei Kreuzvalidierung meist stärker schrumpfen, als nach der Anwendung von Schätzformeln zu erwarten wäre. Zum anderen spielen dabei aber wohl auch Probleme der Schätzformeln selbst (vgl. DARLINGTON, 1968, UHL und EISENBERG, 1970) und Abweichungen der Daten von den Modellannahmen der multiplen Korrelation eine Rolle.

Abgesehen von der Qualität in statistischer Hinsicht ist dann auch schlicht der numerische Umfang einer Stichprobe von Bedeutung, wenn bestimmte, komplexe Analyseverfahren Mindestanforderungen in dieser Hinsicht stellen. Muß man schon bei einer multiplen Regressionsanalyse wenigstens 200 Vpn ansetzen (MARKS, 1966, CLAUDY, 1972), so liegen die Forderungen bei der Kontrastgruppenanalyse (ARMSTRONG und ANDRESS, 1970) und bei der Konfigurationsfrequenzanalyse (LIENERT, 1971) noch wesentlich höher. Ob diese Ansprüche in der vorliegenden Stichprobe erfüllt sind und ob genug Raum für eine Kreuzvalidierung durch Stichprobenteilung bleibt, wird sich im Verlauf der Auswertungen zeigen müssen.

Weitere Begrenztheiten der Aussagen dieser Untersuchung liegen in der Spannweite der Prädiktoren und der Erfassung des Kriteriums. Bei den Prädiktoren fehlen - obwohl die Spannweite größer als bei vielen bisherigen Untersuchungen ist - wichtige Merkmale wie z.B. die in Tests gemessene Intelligenz und Studierfähigkeit. Bei dem hier verwendeten Kriterium "Zensur im Vorexamen", das ohnehin nur einen geringen Teil der interessanten Aspekte des Merkmals Studienerfolg umfaßt, konnten die Studenten, die das Studium abbrachen ohne sich einer Zwischenprüfung zu unterziehen, nicht erfaßt werden. Dadurch dürfte die Variabilität der Fähigkeiten, die bei deutschen Studierenden ohnehin geringer sein dürfte als bei amerikanischen (z.B. TROST, 1975), darüber hinaus weiter eingeschränkt sein.

Insgesamt soll jedoch festgestellt werden, daß der Hauptzweck dieser Studie, verschiedene Analysemodelle in beispielhafter Anwendung gegenüberzustellen, wenig von den Unzulänglichkeiten der Stichprobenerhebung berührt sein dürfte. Dabei sind nämlich vorwiegend allgemeingültige Beobachtungen zu machen, die solange wenig von den spezifischen Eigenschaften der jeweils verwendeten Stichprobe abhängig sind, als in der Stichprobe im Prinzip die gleichen Strukturen bei den Beziehungen zwischen den Merkmalen auftreten wie in der Population. Z.B. wurden ja auch viele Gesetzmäßigkeiten der Allgemeinen Psychologie anscheinend ohne Schaden für den Fortschritt der Wissenschaft an Stichproben gefunden, die nur wenig repräsentativ für die Grundgesamtheit waren, für die die Gesetze gelten sollten.

3. Multiple Regressionsanalyse

Zu Beginn der Ergebniskapitel einige Bemerkungen zum formalen Vorgehen:

Die Arbeit wurde nach der Möglichkeit einer leichteren und übersichtlicheren Darstellung gegliedert; die inhaltlichen und technischen Abhängigkeiten liegen teilweise anders, als es der Aufgliederung der Kapitel und der Abfolge der einzelnen Kapitel entspräche (vgl. 2.2).

Eine Darstellung der Ergebnisse in Form von Tabellen oder graphischen Darstellungen erfolgt nur dann, wenn sie meiner Meinung nach für das Verständnis der Argumentation notwendig oder vom Inhaltlichen her interessant ist. Einzelne, weniger wichtige Ergebnisse, die z.B. Voraussetzungen der verwendeten Verfahren betreffen, werden lediglich in verbaler Form zusammengefaßt, und es werden die Folgerungen gezogen.

Die Merkmale und Studienfächer werden entsprechend den Kurzbezeichnungen in 2.3 benannt. Eine exaktere Definition der Merkmale durch Fragentext und Angabe der technischen Bearbeitung im Verlauf der Datenerfassung und Datenvorbereitung befindet sich in APENBURG et al. (1974).

Alle Berechnungen wurden am Rechenzentrum der Universität Köln und - zu einem relativ geringen Anteil - an einem Tischrechner durchgeführt. Die wichtigeren Programme werden mit Namen und Angabe der Quelle genannt. Ist - bei weniger komplexen Rechnungen - kein Programm angegeben, so wurden einfachere Programme des Pakets SPSS (NIE et al. 1970) am RZ Köln oder ein Tischrechner benutzt.

3.1 Eigenschaften des Verfahrens und notwendige Voraussetzungen der Daten

Die multiple Regressionsanalyse gehört zu den am häufigsten verwendeten multivariaten (multivariat im Sinn der Verwendung mehrerer Variablen) Verfahren überhaupt, und auf dem

Gebiet der Vorhersage muß man sie wohl als das unsere Vorstellungen am meisten prägende Modell bezeichnen. Für die Bevorzugung dieses Modells sind vor allem zwei Gründe verantwortlich zu machen:

- Einmal die Effizienz des Verfahrens, das die Optimierung des Zusammenhangs von relativ vielen Prädiktoren im Verhältnis zum Stichprobenumfang mit einem Kriterium erlaubt.
- Zum anderen die Flexibilität des Verfahrens, das als Modell zwar linear ist, jedoch über Transformationen der Prädiktoren auch kurvilineare Effekte und Wechselwirkungen zu analysieren gestattet. Diese Eigenschaften des Modells wurden allerdings in der Vergangenheit der psychologischen Forschung wohl zu selten in Betracht gezogen (LEE, 1961, DIGMAN 1966).

Als weitere Vorteile sind die gut ausgearbeiteten mathematisch-statistischen Grundlagen, die leichte Verfügbarkeit flexibler Rechenprogramme und die reichen praktischen Erfahrungen mit diesem Modell zu nennen.

Andererseits sind folgende Schwächen des Modells zu nennen:

- Einschränkung auf lineare und additive, sowie durch Variablentransformation faßbare kurvilineare und nichtadditive Effekte
- Schwierigkeit, den Effekt einzelner Prädiktoren innerhalb der Gesamtgleichung zu isolieren, wenn die Prädiktoren untereinander korreliert sind
- Instabilität der Ergebnisse bei kleineren Stichproben und hoher Kollinearität (d.h. hohen Interkorrelationen) der Prädiktoren, wenn auch der Einfluß der Kollinearität durch neuere Rechentechniken reduziert wurde (DRAPER und SMITH, 1966).

Außerdem müssen auf der Seite der Daten bestimmte - teilweise von der Art des zugrundegelegten Populationsmodells abhängige - Modellvoraussetzungen erfüllt sein. Grundsätzlich wird unabhängige Verteilung der Fehler um den Erwartungswert Null mit konstanter Varianz vorausgesetzt. Danach ist bereits eine erwartungstreue und maximal effiziente Schätzung der Parameter

des Modells möglich (z.B. AHRENS, 1968), d.h. die Gewichte der Prädiktoren sind so gewählt, daß eine höchstmögliche multiple Korrelation zwischen Prädiktoren und Kriterium erreicht wird. Sollen zusätzlich statistische Hypothesen z.B. durch F-Tests geprüft werden, so werden Annahmen über die Verteilungsform der Fehlerwerte notwendig. D.h. es wird Normalverteilung der Fehler - in etwas unterschiedlicher Bedeutung je nach dem zugrundegelegten Populationsmodell (z.B. PRESS, 1972) - angenommen, sofern man sich nicht auf angenäherte Gültigkeit des Modells auch ohne Erfüllung der Normalverteilungsvoraussetzung aufgrund des zentralen Grenzwerttheorems verlassen will (vgl. JOHNSTON, 1963). (Die verschiedenen Populationsmodelle werden hier deshalb nicht behandelt, weil ihr Hauptunterschied ähnlich wie bei festen und Zufallsfaktoren bei der Varianzanalyse auf eine unterschiedliche Bewertung der Prädiktoren und damit auf unterschiedliche Inferenzschlüsse hinausläuft, und dieses für die Zwecke der weniger an statistischer Inferenz orientierter Datenanalyse nicht von wesentlicher Bedeutung ist.)

Als letzter Punkt bei den Voraussetzungen ist schließlich die (strittige) Frage der Skalenqualität der beteiligten Merkmale zu nennen. Dabei lassen sich im wesentlichen zwei Richtungen unterscheiden: STEVENS (1951, 1959) als Hauptvertreter der "weak measurement school" (LUBIN, 1962) fordert als Voraussetzung für die Anwendung bestimmter beschreibender (z.B. arithmetisches Mittel, Varianz) und inferenzstatistischer Methoden (z.B. Produkt - Moment - Korrelation, multiple Regressionsanalyse) das Vorliegen von Messungen auf mindestens Intervallskalenniveau. Dieser Standpunkt wird von verschiedenen Methodikern, wie z.B. SIEGEL (1956), in Deutschland LIENERT (1973), propagiert. Im Gegensatz dazu nehmen ebenfalls bekannte und angesehene Methodiker - von LUBIN (1962) als "school of strong statistics" bezeichnet - wie z.B. LORD (1953), ANDERSON (1961), MC NEMAR (1962) und HAYS (1963) den Standpunkt ein, daß statistische Verfahren auf Zahlen und nicht auf Dinge angewendet werden, und daß es insofern kein Verbot der Anwendung bestimmter Verfahren geben kann. Lediglich

die spezifischen Voraussetzungen der jeweils verwendeten Modelle (z.B. Varianzhomogenität, Normalverteilung) sind zu prüfen. Der statistische Test beantwortet die Frage, für deren Beantwortung er entworfen ist, gleichgültig, ob die Messungen stark oder schwach sind (SAVAGE, 1957). Das Problem reduziert sich dann auf die Angemessenheit der auf den Ergebnissen der statischen Verfahren basierenden Schlußfolgerungen (ADAMS, FAGOT und ROBINSON, 1965).

3.2 Die Anwendbarkeit des Verfahrens auf die vorliegenden Daten

3.2.1 Skalenqualität

In bezug auf die Skalenqualität ist vor allem jener - in der Praxis wohl häufigste - Fall problematisch, bei dem das Skalenniveau zwischen Ordinal- und Intervallskala liegt. Man kann solche Daten auch auf einer Intervallskala gemessen auffassen, bei der Zweifel an der Gleichabständigkeit bestehen bzw. gewisse Verzerrungen der Intervalle, d.h. in leichterem Ausmaß nichtlineare Transformatoren vorliegen.

In dieser Situation wird von Vertretern der "weak measurement school", z.B. LIENERT (1973) eine Niveau-Regression vorgeschlagen,d.h. die schwache Intervallskala wird unter Aufgabe von Information als Ordinalskala behandelt, es wird auf die Anwendung parametrischer zugunsten nonparametrischer Verfahren verzichtet.

In der Praxis jedoch wird diese Vorschrift nur sehr selten befolgt; Messungen wie z.B. Schulnoten, gestufte Angaben in Fragebögen, informelle Testskalen werden fast durchweg mit parametrischen Verfahren bearbeitet, wenn vielleicht auch gelegentlich mit schlechtem Gewissen. Man muß sich allerdings fragen, ob dieses in der Praxis übliche Verfahren wirklich so "falsch" ist, oder was nun eigentlich durch die vorgeschlagene Niveau-Regression besser wird. Wenn man zwischen der manifesten (= gemessenen) und der latenten (= "wahren") Variable unterscheidet, so sind

zwar die auf einer schwachen Intervallskala gemessenen Relationen der latenten Variable nicht völlig wirklichkeitsgetreu in der manifesten Variable wiedergegeben. Andererseits ist aber auch nicht einzusehen, wieso die Relationen der latenten Variable bei einer Niveau-Regression auf Ordinalskala in der manifesten Variable besser wiedergegeben werden sollten, bzw. wieso die auf einer Ordinalskala basierenden Schlüsse zutreffender sein sollten (vgl. GAENSSLEN und SCHUBÖ, 1973, S. 37). Insofern scheint mir die vorgeschlagene Niveau-Regression mehr der Beruhigung des Gewissens als der Richtigkeit der Aussagen zu dienen.[1)]

Ein positiver, den Kenntnisstand über die behandelten Fragen erweiternder Ansatz dürfte hingegen der Versuch sein, die Validität des Meßverfahrens nachzuweisen bzw. zu verbessern (LIENERT, 1967, vgl. auch 4.1).

Aufschlüsse über die Wirkung von schwachen Messungen auf bestimmte statistische Verfahren kann man von - bisher anscheinend seltenen - Modelluntersuchungen über die Auswirkungen des in der Praxis zu erwartenden Typs (z.B. Abb. 3.1) von nichtlinearen Transformationen erwarten. BAKER et al. (1967) untersuchten den Einfluß von 35 verschiedenen Typen von Transformationen - wie z.B. in Abb. 3.1 - auf die Ergebnisse von t-Tests. Die Ausgangsdaten waren nach unterschiedlichen Modellen konstruiert worden. Die Übereinstimmung der sich bei den Ausgangsdaten und den transformierten Daten ergebenden t-Werte bei drei als beispielhaft ausgewählten Transformationen zeigt Abb. 3.2. Die Autoren schlossen u.a., daß auch bei sehr strengen Maßstäben die Signifikanzergebnisse der t-Tests von den Transformationen unberührt blieben, sofern nur - vor allem bei stark unterschiedlichen Stichprobenumfängen - zweiseitig getestet wurde. Andererseits zeigten sich bei dem beschreibenden Gebrauch der t-Werte etwas deutlichere Abweichungen der auf den wahren und den transformierten Daten beruhenden Kennwerte, wenn bei den Transformationen die Intervalle nicht in zufälliger Weise verzerrt wurden, sondern wenn die Skalen systematisch in bestimmten Regionen gestaucht und in anderen

---

1) Diese Aussage gilt in erster Linie für den in der empirischen Sozialforschung Tätigen, der vor allem an der Analyse bestimmter, vorliegender Daten interessiert ist. In der experimentellen Psychologie mit ihrer stärkeren Einflußnahme auf die Untersuchungsabsicht gewinnt dagegen der rigorose Standpunkt LIENERTS an Plausibilität.

gestreckt (wie z.B. in Abb. 3.1) wurden. Insgesamt halten die Autoren "starke" statistische Verfahren für mehr als angemessen, mit schwachen Messungen zurechtzukommen, wenn auch bei einzelnen t-Werten - abhängig vom Typ der verzerrenden Transformation - u.U. leichtere Abweichungen auftreten können.

Für die vorliegende Untersuchung wird folgendes angenommen:

1. Grundsätzliche Erwägungen sprechen dafür, statistische Verfahren auf jegliche Daten anzuwenden, für deren Analyse sie formal geeignet sind. Es gibt kein grundsätzliches Verbot der Anwendung bestimmter statistischer Verfahren.

2. Eine Niveau-Regression auf Ordinalskalen-Niveau würde keine Ergebnisse bringen, die den latenten Variablen und ihren Beziehungen zueinander besser entsprechen würden; es würden dadurch jedoch Nachteile bei der Effizienz, Flexibilität und technischen Durchführbarkeit der Datenanalyse entstehen.

3. Dementsprechend wird die Analyse der Daten in der hier vorliegenden Form als die bestmögliche Annäherung an die wahren Verhältnisse genommen, solange keine besseren Meßverfahren z.B. für Schulzensuren oder das Kriterium "Zensur im Zwischenexamen"[1)] bestehen.

4. Die Abweichungen der dabei getroffenen Feststellungen von denen, die man anhand der "wahren" Daten treffen könnte, bleiben ungeklärt. Insgesamt kann man in dieser Hinsicht jedoch eher optimistisch sein, denn die Verzerrungen der manifesten gegenüber der latenten Variablen dürften häufiger in zufälliger als in systematischer Weise erfolgen. So kann nach den Ergebnissen von BAKER et al. (1967) - jedenfalls so weit man die mit Hilfe von t-Tests gefundenen Ergebnisse als hinweisend auch für andere Verfahren betrachtet - auch beim beschreibenden Gebrauch statistischer Kennwerte nur mit geringfügigen Abweichungen gerechnet werden.

Es wird also in der Meßqualität der beteiligten Merkmale kein Hindernis für die Anwendung parametrischer Verfahren gesehen.

---

1) Zu bestimmten Prüfungen des Kriteriums vgl. 4.1

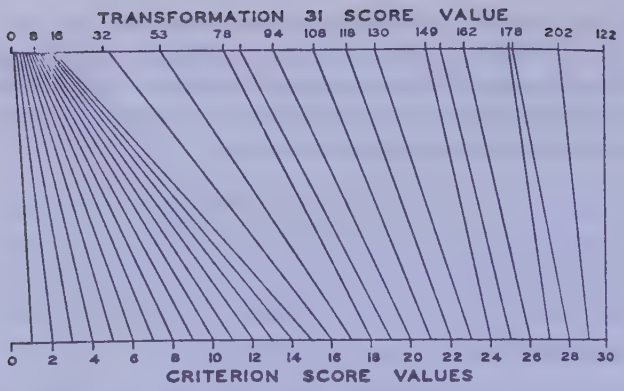


Abbildung 3.1

Beispiel der von BAKER et al. angewendeten Transformationen des "wahren" Merkmals ("Criterion Score Value") in das gemessene Merkmal ("Transformation Score Value") (nach BAKER et al., 1967)

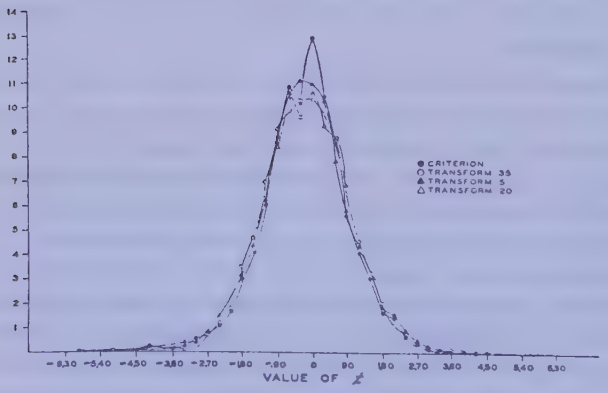


Abbildung 3.2

Übereinstimmung der aus den "wahren" (Symbol •) und aus den transformierten Daten (Symbole ○, △ und ▲) berechneten t-Werte (nach BAKER et al., 1967)

3.2.2 Modellvoraussetzungen

Es soll von Prüfungen berichtet werden, die vor Beginn der eigentlichen Berechnungen angestellt wurden; weitere Tests auf Angemessenheit des Modells sind in 5.3 enthalten. Die hier untersuchten Eigenschaften der Daten sind nur partiell mit den exakten Modellvoraussetzungen identisch. Es ergaben sich die durchgeführten Prüfungen aus dem technisch überhaupt oder mit vertretbarem Aufwand Möglichen. Die Strenge, mit der die Erfüllung der Voraussetzungen untersucht wird, folgt den Richtlinien der meisten Lehrbücher (z.B. JOHNSTON, 1963), wenn auch nicht den stärkeren Ansprüchen von LIENERT (1973) oder BRADLEY (1968).

Zunächst wurde die Verteilungsform aller quanitativen Merkmale untersucht, da diese Verteilungen einen Zusammenhang mit der Verteilung der Residuen aufweisen, wenn auch nicht determinieren. Die Inspektion der Häufigkeitsdarstellungen und der 3. und 4. Verteilungsmomente (Schiefe und Exzeß) erbrachte keine Hinweise auf extreme Abweichungen von der Form der Normalverteilung. Man könnte, um sicher zu gehen, die Verteilungen - wie z.B. von ORLIK (1967) durchgeführt - mit Hilfe einer Flächentransformation nach MC CALL (vgl. LIENERT, 1967) normalisieren. Jedoch waren solche Transformationen, wie z.B. von SCHMIED (1971) berichtet und von mir an anderen, im Gegensatz zu den vorliegenden teilweise extrem schief oder bimodal verteilten, Daten durchgeführt, von zu geringer Wirkung, um den damit verbundenen Aufwand gerechtfertigt erscheinen zu lassen.

Weiterhin wurden die paarweisen Beziehungen der quantitativen Merkmale untersucht. Linearität dieser Beziehungen ist eine Teilvoraussetzung für die Verteilungsannahmen über die Fehlerwerte (vgl. JOHNSTON, 1963). Außerdem waren dabei Hinweise auf mögliche, nützliche Transformationen der Ausgangsmerkmale zu erwarten. Es erwies sich als vom Arbeitsaufwand her noch vertretbar, die jeweils (bei 26 quantitativen Variablen) 325 Paarbeziehungen an den Unterstichpro-

ben der drei größten Studienfächer darzustellen. Von einer Prüfung der Gesamtstichprobe wurde wegen der zu erwartenden Inhomogenitäten durch die Zugehörigkeit der einzelnen Vpn zu den verschiedenen Studienfächern abgesehen. Die lineare Komponente der jeweiligen Paarbeziehung wurde durch das Quadrat des Produkt-Moment-Korrelationskoeffizienten, die gesamte Stärke der jeweiligen Paarbeziehungen einschließlich der linearen und nichtlinearen Komponenten im Eta-Quadrat-Koeffizienten ausgedrückt. Beim Eta-Quadrat-Koeffizienten mußte auch die Richtung des Zusammenhangs beachtet werden, weil dieser Koeffizient nicht symmetrisch wie etwa der Korrelationskoeffizient ist. In der Differenz zwischen r-Quadrat und Eta-Quadrat kommt zum Ausdruck, wieviel Anteil an Varianz durch nichtlineare Komponenten zusätzlich zu der durch die lineare Komponente erklärten Varianz festgelegt werden kann. Welcher Art die nichtlinearen Komponenten sind, bleibt dann anschließend zu klären. Die statistische Prüfung dieses Varianzzuwachses erfolgte mit Hilfe eines F-Tests, wie er auch zum Vergleich zweier multipler Korrelationen verwendet wird (z.B. COHEN, 1968 oder entspr. Lehrbücher). Als Signifikanzgrenze wurde zunächst eine dem Einzeltest zugeordnete Zufallswahrscheinlichkeit von 0,01 festgelegt; dabei wurde jede signifikante Beziehung im Hinblick auf a priori Hypothesen (die z.B. Schicht, Prüfungsangst oder Depression betreffen) oder umfassende Hypothesen (z.B. ob bestimmte Merkmale mit vielen anderen nichtlinearen Beziehungen haben) beurteilt. Zusätzlich wurde ein "experimentbezogenes" Fehlerrisiko nach BONFERRONI (nach KESELMAN, 1974) festgesetzt: Danach gilt als Grenze für den Einzeltest 0,01 : 325 = 0,000031. Die F-Tests wurden näherungsweise durch Vergleich der beobachteten mit kritischen Differenzen durchgeführt, die unter Einsatz verschiedener häufig auftretender fester Größen vorher berechnet worden waren.

Es ergaben sich bei dem Studienfach BWL/VWL 5, Medizin 16 und Math/Inf 15 nach der einzeltestbezogenen Schranke signifikante Zuwächse an erklärter Varianz durch nichtlineare Komponenten. Das ist bei Medizin und Math/Inf etwas häufiger als unter Geltung der Null-

hypothese zu erwarten wäre. Dennoch wurde in keinem Fall die experimentbezogene (Bonnferroni) Schranke der Irrtumswahrscheinlichkeit überschritten, d.h. es waren keine nichtlinearen Komponenten statistisch zu sichern. Die einzeltestbezogen signifikanten nichtlinearen Beziehungen ließen sich nicht in a priori Hypothesen einordnen oder in den Zusammenhang von übergreifenden Hypothesen stellen. Die dabei beteiligten Variablen wurden jedoch für mögliche spätere Querverbindungen protokolliert.

Insgesamt wurden also keine die Anwendbarkeit der multiplen Regressionsanalyse in negativer Weise beeinflussende Schlüsse gezogen.

Abschließend folgen noch zwei Prüfungen, die zwar keine Modellvoraussetzungen betreffen, jedoch am passendsten an dieser Stelle zu nennen sind:

Die von den verschiedenen Zeitpunkten der Haupt- und Nacherhebung (vgl. 2.3.1) stammenden Daten wurden vorwiegend gemeinsam analysiert. Die Berechtigung dieses Vorgehens wurde durch inspektiven und stichprobenweisen Vergleich der Mittelwerte, Varianzen und Kovarianzen aller Merkmale zwischen den drei Personengruppen (Daten vollständig aus der Haupterhebung, Daten aus Haupt- und Nacherhebung, Daten nur aus Nacherhebung, vgl. 2.3.1) untersucht. Dabei ergaben sich keine Hinweise auf stärkere Unterschiede zwischen den Gruppen.

Das zweite Problem betrifft die Behandlung fehlender Angaben. Abgesehen von der Gruppe der nur in der Nacherhebung erfaßten Personen, bei denen eine Reihe von Merkmalen überhaupt nicht vorhanden ist (vgl. 2.3.2), weisen die Daten eine beträchtliche Anzahl von sich zufällig verteilenden fehlenden Angaben auf. Eine einfache und klare Lösung des dadurch entstehenden Problems ist der Ausschluß jeder Vp mit mindestens einer fehlenden Angabe (fallweiser Ausschluß). Da hierbei jedoch viele Informationen verlorengehen und die Stichprobenumfänge stark eingeschränkt werden - die vorliegende Stichprobe würde auf ca. 65 % ihres Ausgangsumfangs reduziert -, wurde vorwiegend eine andere Lösung gewählt. Dabei werden bei der Berechnung von Koeffizienten aus zwei Merkmalen die Vpn mit fehlenden Angaben in einem oder beiden der be-

teteiligten Merkmale jeweils von der Berechnung dieses Koeffizienten ausgeschlossen (paarweiser Ausschluß), so daß bei einer auf diese Weise berechneten Korrelationsmatrix die einzelnen Koeffizienten auf unterschiedlich großen Stichproben basieren. Da die Analyse solcher Korrelationsmatrizen zu Schwierigkeiten führen könnte, wurden bei den drei umfangreichsten Studienfächern die beiden verschiedenen, jeweils auf fall- oder paarweisem Ausschluß der fehlenden Daten beruhenden Datensätze, einem Vorschlag von LÖSEL und WÜSTENDÖRFER (1974) folgend, auf Gleichheit der Mittelwerte, Varianzen und Kovarianzen geprüft. Die Übereinstimmung war sehr gut, so daß keine Zweifel an der Möglichkeit dieser Art der Behandlung fehlender Angaben, soweit sie bei den später angewendeten Verfahren möglich war, entstanden.[1)]

3.3 Ergebnisse

Es werden die Ergebnisse von drei Sätzen von Regressionsanalysen, die jeweils unterschiedliche Gruppen von Prädiktoren verwenden, mitgeteilt. Dabei handelt es sich zum ersten um die Analyse der Abiturnoten als Prädiktoren, zum zweiten um die Analyse der "harten Prädiktoren", d.h. der Merkmale, bei denen die Beurteilung mit nur geringer Wahrscheinlichkeit vom Wissen der Studierenden über ihren Studienerfolg beeinflußt wurde (vgl. 2.3.2), und schließlich um die Analyse "aller" Prädiktoren außer Prüfungsangst und Studiendauer (davon später in 5.3). Eine Aufzählung der in drei Gruppen angeordneten Prädiktoren enthält außer 2.3.2 auch die Tabelle 3.4.

Nachstehend folgen technische Erläuterungen zur Durchführung der Rechnungen und zur Darstellung der Ergebnisse:

---

1) Durch unterschiedliche, durch die Struktur der Verfahren und gelegentlich durch Variablentransformationen erzwungene Techniken der Behandlung fehlender Angaben erklären sich die in den Ergebnistabellen gelegentlich auftretenden, schwankenden bzw. von anderen Tabellen abweichenden Stichprobenumfänge.

- Es wurde das Programm Regressionsanalyse innerhalb des Systems SPSS (NIE et al., 1970) benutzt.

- Das qualitative Merkmal "Vertrauenspersonen" wurde in 5 Dummyvariablen umcodiert (vgl. z.B. COHEN, 1968 oder GAENSSLEN und SCHUBÖ, 1973), die dann einzeln in die Analyse eingingen.

- Die Auswahl der Prädiktoren für die endgültige Regressionsgleichung erfolgte in Stufen: Zunächst für die Abiturnoten allein, dann unter festem Einsatz der Abiturnoten für die "harten Prädiktoren" und schließlich unter festem Einsatz der ausgewählten Abiturnoten und "harten Prädiktoren" für die "weichen Prädiktoren" (vgl. 3.2.2)[1]. Dabei wurde gelegentlich die letzte der in der vorhergehenden Prädiktorengruppe zunächst ausgewählte Variablen wieder weggelassen - wenn sie z.B. von geringer praktischer Bedeutung war -, um die Anzahl zu schätzender Effekte möglichst gering zu halten.

- Die Variablenauswahl erfolgte nach zwei Techniken. Einmal nach dem bekannten "stepwise" Vorgehen, bei dem die durch jeden hinzugefügten Prädiktor zusätzlich erklärte Varianz maximal ist, und eventuell im weiteren Verlauf der Rechnungen überflüssig werdende, früher eingeführte, Prädiktoren ausgeschieden werden. Zum anderen erfolgte die Auswahl der Prädiktoren durch schrittweise Einführung in einer festen, nach Bedeutung und Zeitpunkt der Ausprägung der Merkmale gebildeten, Reihenfolge unter Beibehaltung der Prädiktoren mit wesentlichem Zuwachs an erklärter Varianz. Dieses Verfahren ist technisch aufwendiger, aber in statistischer Hinsicht vorteilhafter, weil eine geringere Anzahl von Entscheidungen anhand der Analysestichprobe getroffen werden muß. Lediglich für die Abiturnoten, bei denen sich keine überzeugende Reihenfolge festlegen läßt, wurde allein das

---

[1] Durch das mehrstufige Vorgehen bleibt die Anzahl zu treffender Entscheidungen geringer, als wenn in jeder Prädiktorengruppe alle Entscheidungen jeweils von Anfang an durchgespielt würden.

"stepwise" Vorgehen gewählt. Beide Verfahren der Variablenauswahl führten im wesentlichen zu den gleichen Ergebnisse, so daß später nur auf das bekannte "stepwise" Verfahren Bezug genommen wird.

- Als statistische Basis für die Entscheidung, wann ein Prädiktor in die Regressionsgleichung aufzunehmen war, diente jeweils ein F-Test auf Verbesserung der Vorhersage durch zusätzliche Prädiktoren (z.B. COHEN, 1968) auf einem Niveau von 0,05. Dieses Risiko ist allerdings nur als heuristisch brauchbar zu werten, da durch das Auswahlverfahren der Prädiktoren an der Analysestichprobe mehrfaches Testen mit Akkumulation des Alpha-Risikos vorliegt (DRAPER et al., 1971). Akzeptiert wurde ein Prädiktor aber nur dann, wenn außer der statistischen ("heuristischen") Signifikanz als weitere Voraussetzung die durch ihn zusätzlich erklärte Kriteriumsvarianz - als Maß für praktische Signifikanz - mindestens 1 % betrug.

  Das gleichzeitige Vorgehen nach statistischer und praktischer Signifikanz wurde gewählt, um einen gewissen Ausgleich zwischen den unterschiedlich großen Stichproben (zwischen ca. 2000 und 100 Vpn) herbeizuführen: Bei großen Stichproben verhindert die Schranke der praktischen Signifikanz, daß viele statistisch signifikante, aber praktisch nutzlose Effekte auftauchen. Bei kleinen Stichproben ermöglicht die großzügig angesetzte statistische Schranke, daß bei allerdings recht hoher Irrtumswahrscheinlichkeit eine Vorhersage in etwa vergleichbarer Höhe erzielt werden kann. Die nominal - vor allem bei Berücksichtigung des erwähnten multiplen Testens - recht hohe Irrtumswahrscheinlichkeit wird dadurch wiederum etwas gemildert, daß für statistische Berechnungen die Zahl der Personen mit vollständigem Datensatz als Stichprobenumfang eingesetzt wird, während die einzelnen Koeffizienten durch den paarweisen Ausschluß von fehlenden Angaben (vgl. 3.2.2) auf höheren Stichprobenumfängen basieren und somit stabiler sein dürften, als im statistischen Test angenommen wird.

- Die Ergebnistabellen 3.1 - 3.3 enthalten die ausgewählten Prädiktoren mit den zugehörigen einfachen und multiplen Korrelationen und Anteilen an erklärter Varianz. Suppressoren sind mit einem Stern kenntlich gemacht. Da vielfach auch die Validitäten der nicht ausgewählten Prädiktoren von Interesse sind, werden in Tab. 3.4 die Validitäten aller Prädiktoren innerhalb der Personenstichproben aufgeführt.

- Die in den Tabellen genannten Personenstichproben beziehen sich auf

  1. die Gesamtheit der befragten Studierenden,

  2. die Studierenden der 9 größten Fächer, bei denen das Kriterium innerhalb jedes Faches auf den Mittelwert 0 und die Streuung 1 transformiert wurde,

  3. die Studierenden der 8 größten Fächer, bei denen außer dem Kriterium auch die quantitativen Prädiktoren fachweise standardisiert wurden,

  4.-11. jeweils die Studierenden der 8 größten Fächer (vgl. 2.3.1).

3.4 Diskussion

3.4.1 Abiturnoten

Die endgültige Beantwortung der Frage, ob sich die Studierenden aller Studienfächer gemeinsam analysieren lassen, wurde zunächst zurückgestellt. Auf jeden Fall erschien schon nach der Betrachtung der Mittelwerte des Kriteriums und der Prädiktoren zwingend, daß mindestens das Kriterium fachweise auf gleichen Mittelwert und Streuung transformiert werden mußte, um unkontrollierte Auswirkungen von Heterogenitäten zwischen den einzelnen Studienfächern zu verhindern. Der Effekt dieser Maßnahme zeigt sich u.a. an den unterschiedlichen Ergebnissen der Personenstichproben 2 und 3 gegenüber 1 in Tabelle 3.1. Durch die Überlegung

Tabelle 3.1: Die Abiturnoten als Prädiktoren

Personenstichproben (Umfang)	Prädiktoren (Abiturnoten)	r	R	Prozent erklärte Varianz	$\hat{R}$ in Population
1. Alle befragten Studierenden (2490)	Deutsch	0,28	0,28	7,8	
	Biologie	0,22	0,31	9,6	
	Englisch	0,22	0,33	10,9	0,326
2. 9 Studienfächer, Kriterium fachweise standardisiert (1786)	Mathematik	0,23	0,23	5,3	
	Deutsch	0,22	0,29	8,4	0,285
3. 8 Studienfächer, Kriterium und Prädiktoren fachweise standardisiert (1307)	Mathematik	0,27	0,27	7,3	
	Deutsch	0,23	0,32	10,2	0,313
4. BWL/VWL (679)	Mathematik	0,31	0,31	9,6	
	Deutsch	0,24	0,35	12,3	0,339
5. Medizin (425)	Deutsch	0,26	0,26	6,8	
	Physik	0,26	0,31	9,6	0,288
6. Mathematik/Inf. (191)	Deutsch	0,28	0,28	7,8	
	Physik	0,25	0,33	10,9	0,283
7. Deutsch (194)	Physik	0,36	0,36	13	
	Musik	0,28	0,42	17,6	
	Erdkunde	0,32	0,45	20,3	0,411
8. Französisch (167)	Erdkunde	0,29	0,29	8,4	
	Deutsch	0,25	0,35	12,3	0,30
9. Chemie (96)	Physik	0,32	0,32	10,2	
	Englisch	0,29	0,35	12,3	0,255
10. Physik (101)	Englisch	0,24	0,24	5,8	0,138
11. Wirtschaftspädagogik (102)	Mathematik	0,38	0,38	14,4	
	Deutsch	0,31	0,44	19,4	0,379

Tabelle 3.2: Die "harten" Prädiktoren (die bereits in Tab. 3.1 ausgewählten Prädiktoren sind durch eckige Klammern kenntlich gemacht)

Personenstichproben (Umfang)	Prädiktoren	r	R	Prozent erklärte Varianz	$\hat{R}$ in Population
1. Alle befragten Studierenden (2490)	[Deutsch	0,26	0,26	7,8	
	[Biologie	0,22	0,29	9,6	
	[Englisch	0,22	0,31	10,9	0,33
2. 9 Studienfächer, Kriterium (1786) standardisiert	[Mathematik	0,23	0,23	5,3	
	[Deutsch	0,22	0,29	8,4	0,29
3. 8 Studienfächer, Kriterium u. Prädiktoren (1307) standardisiert	[Mathematik	0,27	0,27	7,3	
	[Deutsch	0,23	0,32	10,2	0,31
4. BWL/VWL (679)	[Mathematik	0,31	0,31	9,6	
	[Deutsch	0,24	0,35	12,3	0,34
5. Medizin (425)	[Deutsch	0,26	0,26	6,8	
	[Physik	0,26	0,31	9,6	
	Depression	0,11	0,33	10,9	
	Schicht	-0,11	0,35	12,3	0,32
6. Mathematik (191)	[Deutsch	0,28	0,28	7,8	
	[Physik	0,25	0,33	10,9	
	Vertrauenspersonen Geschw.	0,19	0,38	14,4	0,33
7. Deutsch (196)	[Physik	0,36	0,36	13	
	[Musik	0,28	0,42	17,6	
	Depression	0,26	0,48	23	0,45
8. Französisch (167)	[Erdkunde	0,29	0,29	8,4	
	[Deutsch	0,25	0,35	12,3	
	Depression	0,23	0,37	13,7	
	Verhältn.z.d. Lehrern*	0,05*	0,44*	19,4*	0,38
9. Chemie (97)	[Physik	0,32	0,32	10,2	
	Geschlecht	0,27	0,45	20,3	
	Fachinteres.[1]	-0,18	0,49	24	
	Verhältn.z.d. Lehrern	0,27	0,55	30	0,47
10. Physik (101)	[Englisch	0,24	0,24	5,8	
	Vertrauensp.G.	0,33	0,33	10,9	0,23
11. Wirtschaftspädagogik (102)	[Mathematik	0,38	0,38	14,4	
	[Deutsch	0,31	0,44	19,4	
	Geschlecht	0,16	0,48	23	0,41

1) Interesse für Physik
* Suppressorwirkung

Tabelle 3.3: (erste Hälfte)

"Alle Prädiktoren (die bereits in Tab. 3.1 und 3.2 ausgewählten Prädiktoren sind durch eckige Klammern kenntlich gemacht)

Personenstichprobe (Umfang)	Prädiktoren	r	R	Prozent erklärte Varianz	$\hat{R}$ in Population
1. Alle befragten Studierenden (2563)	[Deutsch	0,26	0,26	6,8	
	[Biologie	0,22	0,29	8,4	
	[Englisch	0,22	0,31	9,6	
	Intelligenz	-0,22	0,36	13	0,35
2. 9 Studienfächer, Kriterium fachweise standardisiert (1786)	[Mathematik	0,23	0,23	5,3	
	[Deutsch	0,22	0,29	8,4	
	Intelligenz	-0,21	0,34	11,6	
	Prakt.Begabung	0,14	0,36	13	0,35
3. 8 Studienfächer, Kriterium und Prädiktoren fachweise standardisiert(1307)	[Mathematik	0,27	0,27	7,3	
	[Deutsch	0,23	0,32	10,2	
	Intelligenz	-0,13	0,34	11,6	
	Prakt.Begabung	0,14	0,36	13	0,35
4. BWL/VWL (679)	[Mathematik	0,31	0,31	9,6	
	[Deutsch	0,24	0,35	12,3	
	Mathemat.Beg.	-0,31	0,41	16,8	
	Intelligenz	-0,24	0,44	19,4	
	Prakt.Begabung	0,09	0,45	20,4	
	Kaufmänn.Begab.	-0,11	0,46	21,2	0,44
5. Medizin (425)	[Deutsch	0,26	0,26	6,8	
	[Physik	0,26	0,31	9,6	
	[Depression	0,11	0,33	10,9	
	[Schicht	-0,11	0,35	12,3	
	Prakt.Begabung	0,21	0,42	17,6	
	Intelligenz	-0,17	0,45	20,3	0,42
6. Mathematik/ Informatik (191)	[Deutsch	0,28	0,28	7,8	
	[Physik	0,25	0,33	10,9	
	[Vertrauensp.G.	0,19	0,38	14,4	
	Mathemat.Begab.	-0,22	0,41	16,8	
	Kaufmänn.Begab.	0,18	0,45	20,2	
	Intelligenz	-0,24	0,47	22,1	
	Sprachl.Begab.*	0*	0,49*	24*	0,42
7. Deutsch (196)	[Physik	0,36	0,36	13	
	[Musik	0,28	0,42	17,6	
	[Depression	0,26	0,48	23	
	Prakt.Begabung	0,18	0,52	27	0,48

Tabelle 3.3: (zweite Hälfte)

Personenstichprobe (Umfang)	Prädiktoren	r	R	Prozent erklärte Varianz	$\hat{R}$ in Population
8. Französisch (167)	Erdkunde	0,29	0,29	8,4	
	Deutsch	0,25	0,35	12,3	
	Depression	0,23	0,37	13,7	
	Verhältnis z.d. Lehrern*	0,05*	0,44*	19,4*	0,38
9. Chemie (97)	Physik	0,32	0,32	10,2	
	Geschlecht	0,27	0,45	20,3	
	Fachinteresse[1]	-0,18	0,49	24	
	Verhältnis z.d. Lehrern	0,27	0,55	30	
	Intelligenz	-0,37	0,59	34,8	0,51
10. Physik (101)	Englisch	0,24	0,24	5,8	
	Vertrauenspersonen Geschwister	0,15	0,33	10,9	0,23
11. Wirtschaftspädagogik (102)	Mathematik	0,38	0,38	14,4	
	Deutsch	0,31	0,44	19,4	
	Geschlecht	0,16	0,48	23	
	Praktische Beg.	0,20	0,53	28	0,45

1) Interesse für Physik

* Suppressorwirkung

Tabelle 3.4: Die Validitäten aller Prädiktoren

	Personenstichproben										
	1	2	3	BWL 4	Med 5	Math 6	Deu 7	Fra 8	Chem 9	Phys 10	WiPä 11
"harte Prädiktoren"											
Deutsch	0,28	0,22	0,23	0,24	0,26	0,28	0,10	0,25	0,20	0,20	0,31
Englisch/1.Fr.	0,24	0,19	0,20	0,19	0,18	0,24	0,14	0,16	0,29	0,25	0,27
Erdkunde	0,21	0,19	0,21	0,17	0,23	0,22	0,32	0,29	0,14	0,06	0,21
Mathematik	0,17	0,23	0,28	0,31	0,25	0,16	0,27	0,20	0,21	0,15	0,38
Physik	0,21	0,21	0,24	0,19	0,26	0,25	0,36	0,22	0,32	0,18	0,22
Biologie	0,23	0,17	0,13	0,16	0,22	0,13	0,20	0,22	0,27	0,05	0,20
Musik	0,13	0,10	0,04	0,07	0,11	0,13	0,28	0,16	0	-0,03	0
Geschlecht	-0,03	0,03	0,04	0	0,05	0,05	-0,09	0,01	0,27	0	0,16
Schicht	0,04	0	-0,02	0,04	-0,11	0,11	-0,06	0	-0,19	0,22	-0,11
Depression	0,10	0,10	0,10	0,08	0,11	0,09	0,26	0,08	-0,06	-0,03	0,20
Verh.z.d.Lehr.	0,08	0	0,08	0,02	0,09	-0,04	0,14	0,06	0,27	0	0,17
Fachinteresse	-	-	-	-	-0,10	-0,04	0,08	0,09	-0,18	0,05	-
Vertrauensp.*	-	-	-	-	-	-	-	-	-	-	-
"weiche Prädiktoren"											
Intelligenz	-0,22	-0,21	-0,13	-0,24	-0,17	-0,24	-0,18	-0,14	-0,37	-0,10	-0,16
Mathem.Begab.	-0,06	-0,18	-0,18	-0,31	-0,16	-0,22	-0,16	-0,17	-0,25	0	-0,24
Sprachl.Begab.	-0,08	0	0	0	0	0	-0,15	0,04	-0,15	-0,10	0,20
Naturw. Begab.	-0,09	-0,08	-0,09	-0,09	-0,20	-0,15	0	0,07	-0,22	0,10	0
Künstl. Begab.	-0,02	0,04	0,04	0,08	0,05	0,05	-0,14	0,08	-0,08	0	0,08
Techn. Begab.	0,06	0,04	0,05	0,03	0,08	0,02	0,11	-0,04	0,02	0,13	0,09
Kaufm. Begab.	0,15	0,03	0,04	-0,11	0,08	0,18	0,17	-0,09	0,02	0,04	0,13
Prakt. Begab.	0,13	0,14	0,14	0,10	0,21	0,17	0,18	0,08	0,16	0,10	0,20
Leistungsmot. im Selbstbild:											
Hoffn.a.Erf.	-0,11	-0,07	-0,08	-0,06	-0,08	-0,20	-0,14	-0,08	-0,17	0,08	-0,04
Furcht v.Miß.	0,05	-0,02	0,03	0	0,10	0,03	0,12	0,05	-0,12	0,09	-0,08
Neutral	0	0	0	0	0,09	-0,07	0,17	-0,04	-0,10	-0,08	0

* keine Angaben, weil in 5 Dummy-Variablen aufgelöst

und die Feststellung, daß auch bestimmte Prädiktoren systematisch in den Studienfächern verschieden sind - z.B. durch den n.c. in Medizin -, war die entsprechende Standardisierung bei der Personenstichprobe 3 nahegelegt worden. Hier müssen dann auch die Prädiktorausprägungen fachbezogen interpretiert werden, d.h. gute Abiturzensuren bedeuten dann z.B. jeweils gute Zensuren im Verhältnis zum Durchschnitt der Studenten in dem betr. Studienfach. Inhaltlich scheint mir diese Standardisierung auch der Prädiktoren durchaus vertretbar und aus theoretischen Gründen sogar vorzuziehen zu sein, wenn auch die Auswirkung dieser Maßnahme auf die Ergebnisse der Tabellen 3.1 - 3.4 nicht übermäßig groß ist: Beide Personenstichproben 2 und 3 weisen die gleichen Prädiktorenauswahlen auf, und die erklärten Varianzteile und die einzelnen Validitäten unterscheiden sich nicht sehr wesentlich.

Die fachweise Standardisierung aller beteiligten Variablen muß allerdings keineswegs auch Unabhängigkeit der Fehlerwerte bewirken. Vielmehr könnten immer noch durch die Abweichung der in der Gesamtgruppe entwickelten Regressionsgleichung von der Gleichung mit optimaler Anpassung an das jeweilige Fach die Fehler innerhalb der Fächer korreliert sein (vgl. 5.3).

Die Höhe der in Tab. 3.1 dargestellten multiplen Korrelationen liegt zwischen 0,29 und 0,33 für die Gesamtgruppen (1-3) und zwischen 0,31 und 0,45 für die einzelnen Studienfächer. Die Größenordnung ist der von anderen deutschen Untersuchungen vergleichbar, ist aber - wie bereits in 2.1 genannt - aus einleuchtenden Gründen niedriger als in den USA. Zur Vorhersage im Einzelfall liegen die Zusammenhänge - wie schon oft konstatiert wurde - sicherlich zu niedrig.

Für den Vergleich der Personenstichproben ist zu bedenken, daß innerhalb von Stichproben bestimmte multiple Korrelationen eine - mit kleinerem Stichprobenumfang steigende - systematische Überschätzung der "wahren", in der Grundgesamtheit zu erwartenden multiplen Korrelationen darstellen. Deshalb sind in den letzten

Spalten der Tab. 3.1 - 3.3 die in der Grundgesamtheit unter Berücksichtigung des Stichprobenumfangs zu erwartenden multiplen Korrelationen angegeben. Es ist zu beachten, daß die übliche Schrumpfungskorrektur nach WHERRY (z.B. GUILFORD, 1965, S. 401, GAENSSLEN und SCHUBÖ, 1973, S. 114, McNEMAR, 1962) unangemessen ist; sie schätzt die Validität der in der Population bestimmten Gleichung in einer Zufallsstichprobe (DARLINGTON, 1968, UHL und EISENBERG, 1970). Nach dieser Formel wird die zu erwartende Schrumpfung zu niedrig angegeben, was möglicherweise in Feststellungen, z.B. von COOLEY und LOHNES (1971), daß die in Kreuzvalidierungen gefundene Schrumpfung häufig größer als die erwartete ist, seinen Niederschlag findet. Eine angemessenere Schätzung gibt die unabhängig von LORD (1954) und NICHOLSON (1960, beide Arbeiten zitiert nach DARLINGTON, 1968) entwickelte Formel, die von STEIN (1960, nach DARLINGTON, 1968) unter Berücksichtigung realistischerer Annahmen modifiziert wurde. Nach dieser "Formel 14" bei DARLINGTON (1968) wurden die in den Tabellen aufgeführten Schätzwerte bestimmt. Dabei wird allerdings nicht berücksichtigt, daß durch die Prädiktorenauswahl an der Analysestichprobe eine zusätzliche, schwer abzuschätzende Schrumpfung zu erwarten ist. Dieses Risiko soll im Sinn eines heuristischen Vorgehens akzeptiert werden.

Bei Vergleich der jeweils in der Population zu erwartenden multiplen Korrelationen liegt die Gesamtgruppe (2 und 3) etwa in der Mitte, weil hier die erwartete Schrumpfung sehr gering ist. Die höchsten Zusammenhänge zeigen die Studienfächer Deutsch und Wirtschaftspädagogik, die beide die höchsten Anteile vom Lehramtstudenten aufweisen. Eher niedrig bzw. am niedrigsten liegen Physik, Chemie, Medizin und Mathematik/Informatik, die Studiengänge, deren Inhalte von der schulischen Ausbildung recht weit entfernt und deren Studenten relativ stark ausgelesen sind. Ob diese Eigenschaften der Studiengänge ursächlich für die schlechtere bzw. bessere Vorhersagbarkeit sind, kann hier nicht entschieden werden, obgleich eine gewisse Plausibilität dafür spricht.

Die Auswahl der einzelnen Prädiktoren in Tab. 3.1 muß unter Berücksichtigung der Validitäten aller Prädiktoren beurteilt werden. Oft entscheiden nämlich nur geringe Unterschiede der Validitäten bei den Prädiktoren über Auswahl oder Nichtauswahl eines Prädiktors, so daß teilweise auch eine etwas andere Prädiktorenauswahl zu nahezu gleich hoher multipler Korrelation hätte führen können. Darin kommt, ebenso wie sich auch beim Verfolgen der partiellen Korrelationen der noch nicht in die Regressionsgleichung eingeführten Prädiktoren mit dem Kriterium im Verlauf der schrittweisen Einführung der Prädiktoren bestätigen läßt, eine recht hohe Multikollinearität der Prädiktoren zum Ausdruck. Eine exakte Analyse und größenmäßige Erfassung der Multikollinearität z.B. nach der Methode von FARRAR und GLAUBER (1967) wäre nur zur Planung einer neuen Datenerhebung notwendig und von größerem Nutzen (JOHNSTON, 1963). Sie wurde hierdeswegen nicht durchgeführt. Ebenfalls hätte - jedenfalls bei den Stichproben mit deutlich weniger als 200 Vpn - wohl auch ein einfacheres, z.B. den einzelnen Validitäten folgendes Gewichtungsverfahren zu etwa gleich hohen, in der Population zu erwartenden "multiplen" Korrelationen führen können (vgl. MARKS, 1966, CLAUDY, 1972). Jedoch wurde hier die Methode der Regressionsanalyse wegen ihrer Flexibilität z.B. in bezug auf Suppressoreffekte (vgl. GAENSSLEN und SCHUBÖ, 1973, S. 126) und wegen der leichteren Handhabung beibehalten. In Folge der zwischen den Prädiktoren bestehenden Multikollinearität erklärt der erste ausgewählte Prädiktor jeweils auch den größten Teil der überhaupt erklärbaren Kriteriumsvarianz, weil er die mit den anderen Prädiktoren gemeinsame Kriteriumsvarianz voll auf sich selbst vereinigt.

Es scheint eine gewisse Tendenz zu bestehen, daß die wichtigsten Abiturfächer Deutsch und Mathematik sowohl bei der gemeinsamen Analyse verschiedener Studienfächer (Stichproben 2 und 3) als auch bei den einzelnen Studienfächern häufiger als andere Fächer von besonderer Bedeutung für die Vorhersage des Studienerfolgs sind. Relativ selten treten Musik (beim Studienfach Deutsch), Biologie (beim Studienfach Chemie) und Erdkunde (beim Studienfach Deutsch) auf.

Auffällig ist, daß bei den Studienfächern Math/Inf, Deutsch, Französisch und Physik die entsprechenden Schulfächer, und bei Medizin das am nächsten verwandte Schulfach Biologie, keineswegs die höchsten oder besonders hohe Validitäten aufweisen und auch nicht als Prädiktoren ausgewählt wurden. Z.B. lauten die höchsten, der Grösse nach geordneten Validitäten für Math/Inf:

Deutsch	0,28
Physik	0,25
Englisch	0,24
Erdkunde	0,22
Mathematik	0,16 usw.

ebenso ergeben sich für das Studienfach Deutsch:

Physik	0,36
Erdkunde	0,32
Musik	0,28
Mathematik	0,27
usw. ... bis	
Deutsch	0,10 (mit der niedrigsten Validität überhaupt).

Offensichtlich weisen also die dem Studienfach homologen oder am nächsten stehenden Schulfächer eher eine geringe Validität und Bedeutung für die Vorhersage der Zensur des Zwischenexamens auf, so daß teilweise sogar mit dem Studienfach recht wenig verwandte Schulfächer die besten Vorhersageeffekte erzielen.

Dieses frappante Ergebnis ist nicht vollständig neu; zumindest konnte man anderen Untersuchungen entnehmen (z.B. KOPFMANN, 1969; WEINGARDT, 1969), daß die Abiturzensur im homologen Schulfach keineswegs die hervorstechend hohe Validität aufwies, die man eigentlich erwarten sollte, und Weingardt berichtet von schwachen Tendenzen in wenigen, kleineren Untersuchungen, daß gerade das homologe Schulfach eine besonders niedrige Validität aufweist. Schmied (1971) schließlich fand bei Studierenden für das Lehramt an Gymnasien eine von nur wenigen Ausnahmen durchbrochene Tendenz, daß

sich die Abiturnoten des korrespondierenden Schulfaches mit ihrer Validität nur im Mittelfeld aller Abiturfächer befinden.

Zur Klärung dieses Faktums kann man vermuten, daß die Streuung der Zensuren im homologen Schulfach durch Selbstselektion (BINDER-KRAUTHOFF, 1971), teilweise auch durch Selektion von außen (KOPFMANN, 1969) eingeschränkt ist und daß dadurch, weil ja die Höhe des Korrelationskoeffizienten von der Streuung der beteiligten Merkmale abhängt, die Validität nicht in ihrer eigentlichen Höhe gemessen wird. In diesem Fall läßt sich, wenn man das Verhältnis der Streuungen in der Grundgesamtheit aller Abiturienten und in der Stichprobe, bestehend aus den Studenten des betreffenden Studienfachs, kennt, die Höhe des Validitätskoeffizienten abschätzen, die er bei nicht eingeschränkter Streuung erreicht hätte.

Zur Prüfung dieser Frage sind in Tabelle 3.5 Mittelwerte und Streuungen der Abiturzensuren für die Studierenden des entsprechenden Studienfaches und für die Gesamtstichprobe aufgeführt. Im unteren Teil der Tabelle handelt es sich um das dem Studienfach Medizin am nächsten verwandte Schulfach Biologie. Beim Studienfach Französisch ist es unklar, ob mit der ersten Fremdsprache bei den einzelnen Studierenden Englisch oder Französisch gemeint war.

Tabelle 3.5: Mittelwerte, Streuungen und Stichprobenumfänge für Abiturnoten von Studierenden in entsprechenden Studienfächern und für die Gesamtstichprobe der Kandidaten

	ABITURNOTEN der					
	Studierenden in entspr. Studienfächern			Gesamtstichprobe der Kandidaten		
SCHULFÄCHER	$\bar{X}$	S	N	$\bar{X}$	S	N
Mathematik	1,71	0,74	258	2,86	1,1	2492
Deutsch	2,38	0,67	193	2,94	0,84	2492
Engl./1.Fremdsprache	Studienf.: Franz. 2,33	0,8	166	3,03	0,94	2487
Physik	1,96	0,71	101	2,73	0,91	2468
Biologie	Studienf.: Medizin 2,04	0,82	415	2,46	0,82	2385

Die Tabelle zeigt, daß die Abiturnoten der schulfachkorrespondierenden Studienfächer und der Gesamtstichprobe die vermuteten Mittelwerts- und Streuungsunterschiede aufweisen. Unter Verwendung der Angaben aus Tab. 3.5 wurden die Validitäten aufgewertet (vgl. GUILFORD, 1965, S. 343). Danach steigen die Vailiditäten auf folgende Werte:

Mathematiknote (für das Studienfach Mathematik)
von .16 auf .23

Deutschnote (für das Studienfach Deutsch)
von .10 auf .12

Engl./Franz.-Note (für das Studienfach Franz.)
von .16 auf .19

Physiknote (für das Studienfach Physik)
von .18 auf .23.

Die Ergebnisse der Aufwertungsrechnungen zeigen, daß bei einer Berücksichtigung der eingeschränkten Streuungen die Validitäten etwas ansteigen, jedoch werden die Noten der relevanten Schulfächer dadurch keineswegs zu den stärksten Prädiktoren.

Zur Erklärung bleiben jetzt noch die zahlenmäßig beträchtlichen Mittelwertsunterschiede. Die mögliche Wirkung dieses Unterschieds sei an einem Beispiel erläutert:
Beim Wettkampf einer stark nach Leistung ausgelesenen und damit leistungshomogenen Gruppe von Sportlern - etwa bei einer Olympiade - werden für das gute Abschneiden weniger die sportliche Leistungsfähigkeit als vielmehr Randfaktoren wie Konzentrationsfähigkeit, Streßanfälligkeit und ähnliche Eigenschaften bestimmend sein. Das würde heißen, bei den Studenten der schulfachkorrespondierenden Studienfächer handelt es sich um eine in bezug auf die Leistungsfähigkeit im entsprechenden Schulfach recht stark ausgelesene Gruppe, und für den Examenserfolg sind dann etwas weiter abliegende Merkmale stärker bestimmend als die eigentlich wichtigen Merkmale. Daß sich dieser Sachverhalt in einer nur wenig verringerten Streuung der entsprechenden Abiturnote ausdrückt (s. Tab. 3.5), könnte an einem größeren Anteil fehlerbedingter Varianz im Bereich guter Zensuren liegen.

Die Ursache dafür könnte sein, daß unterschiedliche Zensurmaßstäbe der Lehrer sich besonders bei extremen Noten auswirken - ein Lehrer gibt kaum jemals eine 1, ein anderer viel bereitwilliger - und daß die Zuverlässigkeit der Differenzierung am oberen Ende der Notenskala geringer ist, weil in Klassenarbeiten und anderen Prüfungen ein mittlerer Schwierigkeitsgrad vorherrscht.

Andererseits könnte aber auch, selbst wenn gute Zensuren gleich reliabel wie mittlere wären, die Korrelationen im Bereich der guten Zensuren mit dem Kriterium niedriger sein als im mittleren Bereich. Schließlich ist denkbar, daß mehrere dieser Ursachen zusammenwirken und gemeinsam zu einer geringeren prognostischen Gültigkeit der eigentlich relevanten Merkmale führen.

Die in Tab. 3.5 dargestellten Unterschiede zwischen den Fächern, d.h. zwischen jeweils einem Studienfach und allen anderen, illustrieren überzeugend, daß - wie eingangs dieses Abschnitts (3.4.1) ausgeführt - zur gemeinsamen Analyse der Studienfächer eine fachweise Standardisierung der Variablen zur Ausschaltung der Mittelwerts- und Streuungsunterschiede notwendig war. Darüberhinaus zeigt sich an dem Zusammenhang der Abiturnoten nicht nur mit dem Examenserfolg, sondern auch mit der gewählten Fachrichtung, daß die Abiturnoten ebenso gut auch als Prädiktoren des gewählten Studienfachs aufgefaßt werden können. Dieses Faktum, dem in Deutschland in Zusammenhang mit dem Studienerfolg bisher lediglich von ORLIK (1967) Aufmerksamkeit gezollt wurde, soll u.a. in Kap. 4 untersucht werden.

3.4.2 Die "harten" Prädiktoren

Die hierbei wichtigste Frage dürfte sein, ob die den Abiturnoten hinzugefügten Prädiktoren Geschlecht, Schicht, Depression, Verhältnis zu den Lehrern, Fachinteresse und Vertrauensperson überhaupt zusätzlich weitere Kriteriumsvarianz erklären können. Bei den gemeinsamen Analysen der Studienfächer ergeben sich keine solchen Effekte, offensichtlich existieren hier keine fachübergreifenden Tendenzen wie bei den Abiturnoten. Auch bei BWL/VWL bleiben die Abiturnoten in einer Art Basisvorhersage allein. Für alle anderen Studienfächer läßt sich die Vorhersagegenauigkeit durch die hinzugefügten Merkmale steigern: relativ wenig bei Medizin, Math/Inf und Wirtschaftspädagogik, recht beträchtlich bei Chemie mit einer Steigerung der erklärten Varianz von ca. 10 % auf 30 %. Dabei sind allerdings die zu erwartenden Schrumpfungen, die ja besonders bei den kleinen Stichproben ins Gewicht fallen, noch nicht berücksichtigt. Insgesamt bleiben die Abiturnoten deutlich die wichtigsten Prädiktoren.

Welche der hinzugefügten Prädiktoren einen Effekt erzeugen, schwankt stark von Studienfach zu Studienfach; nachstehend folgen Bemerkungen zu einzelnen Effekten:

Persönlichkeit: Übereinstimmend mit amerikanischen Befunden treten Merkmale dieses Bereichs (Depression und Vertrauenspersonen) recht häufig als Prädiktoren auf. Danach erreichen Studenten der Medizin, Deutsch und Französisch mit niedriger Depression bessere Examenszensuren. Studenten der Math/Inf und Physik mit Geschwistern als Vertrauenspersonen erzielten eher schlechtere Examenszensuren. Der in bezug auf die Studenten der Math/Inf und Physik schwer einsehbare Befund läßt sich zumindest für Math/Inf in Zusammenhang mit Ergebnissen von AMELANG (1971) und SCHLATTMANN und APENBURG (1976) sehen, daß Mathematikstudenten im Durchschnitt introvertierter als andere Studenten sind und dann besser mit dem Studium zurechtkommen,

wenn sie diesem Durchschnitt entsprechen. Für die vorliegenden Befunde würde das bedeuten, daß Studenten der Math/Inf und vielleicht auch der Physik mit Geschwistern als Vertrauenspersonen weniger introvertiert sind und insofern nicht so gut mit dem Studium zurechtkommen; in der Tendenz schnitten Studierende dieser Fächer ohne Vertrauenspersonen eher besser im Examen ab.

Geschlecht: Studentinnen der Chemie legen schlechtere Zwischenexamina ab als Studenten dieses Fachs. In den USA gilt umgekehrt, daß Studentinnen bessere Zensuren während des Studiums erzielen, vermutlich weil sie einerseits durch geringeren äußeren Druck höher intrinsisch motiviert sind und andererseits mit Eigenschaften wie Anpassung, Gefügigkeit, Passivität und geringerer Selbständigkeit günstigere Voraussetzungen für gute Zensuren mitbringen. In Deutschland sind die Befunde uneinheitlich, jedoch werden eher schlechtere Studienzensuren und insgesamt weniger erfolgreiches Studieren der Studentinnen angenommen (KATH et al., 1966, EHLERS und GIESEN, 1971). Erstaunlicherweise haben jedoch Mädchen eher bessere Zensuren auf der Oberschule (BARTENWERFER und GIESEN, 1973, APENBURG et al., 1974, ORLIK, 1967). Vielleicht stehen diese Erscheinungen im Zusammenhang mit der von BECKMANN et al. (1972) gefundenen Entwicklung der Studentinnen während ihres Studiums in Richtung einer sich verschlechternden Integration in Hochschule und Studium.

Schicht: Zu diesem Merkmal werden sowohl aus den USA als auch aus Deutschland heterogene Befunde berichtet; gelegentlich wurden U-förmige Zusammenhänge entdeckt. Der hier gefundene lineare Effekt, daß Medizinstudenten aus unterer Schicht bessere Physikumszensuren haben, erscheint zunächst erstaunlich. Er muß im Zusammenhang mit der Situation in den anderen Studienfächern gesehen werden, in denen die - wie bekannt - schlechteren Voraussetzungen der Angehörigen unterer Schichten für ein erfolgreiches Durchlaufen des Bildungssystems durch solche (Selbst-) Ausleseprozesse kompensiert werden, daß sich keine Zusammenhänge zwischen sozialem Status und Schul- oder Examenszensuren ergeben. Wegen der speziellen Situation im Studienfach Medizin - wenigstens in der

Vergangenheit Privileg höherer Schichten, sowie durch den n.c. vielfach erzwungene Wartezeit vor dem Studium - dürften Angehörige unterer Schichten in der Regel nur dann Medizin studieren, wenn sie glänzende Abiturzeugnisse und (oder) eine starke Motivation zum Medizinstudium haben (BOCHNIK, 1974), wodurch wiederum das bessere Abschneiden dieser Studenten im Physikum plausibel gemacht wird.

Fachinteresse: Dieses nur für einzelne Studienfächer definierte Merkmal brachte insgesamt kaum Effekte, obwohl in den USA sehr starke Prädiktoren gerade aus dem Interessenbereich gefunden wurden. Ursächlich für die fehlende Wirkung dieses Bereichs dürfte wohl die starke Auslese der deutschen Studenten sein; die Übereinstimmung der Interessen mit dem gewählten Studienfach ist gleichsam eine Selbstverständlichkeit. Der Prädiktoreffekt dieses Merkmals im Studienfach Chemie könnte damit zusammenhängen, daß es sich hier - in leichter Abweichung von der Definition - nicht um Interesse für das homologe Schulfach Chemie, sondern um das verwandte Fach Physik handelt, und daß die Chemiestudenten in bezug auf das Interesse für Physik möglicherweise weniger stark ausgelesen sind. Parallel dazu weist bei den Chemiestudenten die Abiturnote in Physik die höchste Validität aller Abiturnoten auf.

Verhältnis zu den Lehrern: Interessant ist der bei den Angehörigen des Studienfachs Französisch auftretende Suppressoreffekt des Merkmals "Verhältnis zu den Lehrern"[1]: Studierende mit gutem Verhältnis zu den Lehrern erreichten ein schlechteres, Studierende mit schlechtem Verhältnis zu den Lehrern ein besseres Zwischenexamen als aus der Kenntnis der Abiturnoten allein vorhergesagt wird. Es wird anscheinend die in den Zensuren enthaltene, von der "wirklichen Leistung" des Schülers wegführende Wechselbeziehung

---

1) Die Bestimmung derjenigen Variablen, aus denen der Suppressor irrelevante Varianz entfernt, geschah durch Verfolgen der partiellen Korrelationen des Suppressors mit dem Kriterium im Verlauf der schrittweisen Auspartialisierung der anderen ausgewählten Prädiktoren aus dem Kriterium.

zwischen Schüler und Lehrer (z.B. Voreingenommenheit des Lehrers bei der Beurteilung, Auswirkung der persönlichen Beziehung auf die Motivation des Schülers) durch diesen Suppressor korrigiert. Bei den Fächern Math/Inf und Physik war dieser Effekt in der Tendenz ebenfalls zu beobachten, bei BWL/VWL lag er knapp oberhalb der Grenze der Zufallswahrscheinlichkeit.

Das Studienfach Chemie weist erstaunlicherweise einen direkten Prädiktionseffekt dieses Merkmals auf, d.h. Studenten mit gutem Verhältnis zu den Lehrern haben eher schlechtere Examensnoten, und umgekehrt. Mit einiger Vorsicht könnte man vermuten, daß wegen der geringen Repräsentation des Fachs Chemie in der Schule chemiebegeisterte Schüler, die später erfolgreich Chemie studieren, während der Schulzeit nicht so gute Beziehungen zu den meisten Lehrern haben.

3.4.3 "Alle" Prädiktoren

Hier wurden den "harten" die "weichen" Prädiktoren angefügt. Es handelt sich dabei um Merkmale der Begabung (allgemeine, sprachliche, mathematische, naturwissenschaftliche, künstlerische, technische, kaufmännische und praktische Begabung), die möglicherweise, weil sie in der Selbstbeurteilung der Studenten erhoben wurden, vom Wissen der Studenten von ihrem Studienerfolg quasi rückwirkend beeinflußt wurden und deshalb als "weiche Prädiktoren" in einer eigenen Gruppe zusammengefaßt wurden.

Es finden sich Effekte, d.h. allgemeine Tendenzen der Studienfächer in den Gesamtgruppen (2 und 3) bei allgemeiner geistiger und praktischer Begabung. Eine hohe Meinung von der eigenen geistigen und eine niedrige Meinung von der eigenen praktischen Begabung sagen gute Zensuren im Zwischenexamen voraus. Diese Tendenzen unterstreichen die Theoriebezogenheit der meisten Hochschulstudiengänge.

Bei den einzelnen Studienfächern sind die Ergebnisse wechselnd; häufig findet man - wie zu erwarten - Intelligenz und praktische Begabung. Im Einzelnen fällt auf:

- Beim Studienfach BWL/VWL ist "kaufmännische Begabung" sinngemäß positiv, bei Math/Inf negativ mit dem Erfolg im Vordiplom korreliert. U.a. darin kommen klare Unterschiede zwischen den Studienfächern zum Ausdruck.

- Bei Math/Inf tritt "sprachliche Begabung" als Suppressor zu Abiturnoten und allgemeiner Intelligenz[1)] auf. Anscheinend werden dadurch irrelevante sprachliche Komponenten aus den Merkmalen des Schulerfolgs und der allgemeinen Intelligenz entfernt.

- Bei den Studienfächern Französisch und Physik ergeben sich keinerlei Effekte durch "weiche Prädiktoren", während bei allen anderen Personenstichproben die Vorhersagegenauigkeit durch "weiche Prädiktoren" gesteigert werden konnte. Physik bleibt dadurch das Fach mit einer relativ und absolut sehr geringen Vorhersagbarkeit des Kriteriums.

3.5 Schlußfolgerungen

Grundsätzliche Überlegungen bezüglich der Meßqualität der beteiligten Merkmale und empirische Überprüfungen der Modellannahmen erbrachten keine Hindernisse für die Anwendbarkeit der multiplen Regressionsanalyse als dem im vorliegenden Problembereich üblichen Standardverfahren. Inhaltlich wurden Aufschlüsse über die Wirkung eines über die Abiturnoten hinaus erweiterten Prädiktorenspektrums erzielt, und es ergaben sich Hinweise darauf, ob eine gemeinsame Analyse der verschiedenen Studienfächer möglich ist.

---

1) Die Bestimmung derjenigen Variablen, aus denen der Suppressor irrelevante Varianz entfernt, geschah durch das Verfolgen der partiellen Korrelationen des Suppressors mit dem Kriterium im Verlauf der schrittweisen Auspartialisierung der anderen ausgewählten Prädiktoren aus dem Kriterium.

Es erwies sich, daß eine fachweise Transformation des Kriteriums und wohl auch der Prädiktoren auf gleichen Mittelwert und Streuung zur gemeinsamen Analyse der Studienfächer notwendig war, um entsprechende Heterogenitäten zwischen den Fächern auszuschalten. Dabei hatte die Transformation auch der Prädiktoren nur geringe Auswirkung auf die Ergebnisse der Regressionsanalysen, jedoch dürfte dieses Verfahren aus prinzipiellen Gründen wohl vorzuziehen sein. Trotz der durchgeführten Transformationen bleiben Zweifel bestehen, ob die Voraussetzung unkorrelierter Fehler für die Gesamtgruppen erfüllt sind. Die endgültige Untersuchung dieser Frage soll bis 5.3 zurückgestellt werden.

Allgemeine, in etwa allen Studienfächern innewohnenden Tendenzen bei der Vorhersage des Kriteriums ergaben sich vor allem bei den Abiturnoten Deutsch und Mathematik, weiterhin auch bei den "weichen Prädiktoren" Selbstbeurteilung der allgemeinen geistigen und der praktischen Begabung.

Bei den einzelnen Studienfächern zeigten sich Vorhersageeffekte durch die Abiturnoten in der erwarteten Höhe. Diese Vorhersage ließ sich teilweise durch Hinzufügung weiterer Prädiktoren, steigern. Dennoch bleiben die erzielten multiplen Korrelationen zu gering für eine Vorhersage im Einzelfall, d.h. der Wunsch nach einer genaueren Vorhersage eventuell durch besser passende methodische Verfahren bleibt bestehen.

Es wurden mehrere, einleuchtend interpretierbare Suppressoreffekte beobachtet, durch die die Vorhersage verbessert werden konnte; sie traten allerdings - wie auch in der Literatur berichtet wird - an schwer vorherzusagender Stelle auf, d.h. eine systematische Suche nach Suppressoren erscheint nach wie vor problematisch.

In bezug auf praktische Anwendung kommen die meisten der hier zusätzlich zu den Abiturnoten erprobten Prädiktoren höchstens für Beratungs-, kaum aber für Auslesemaßnahmen in Frage. Ein interessanter Aspekt in bezug auf Anwendung in der Auslese tritt allerdings bei dem Merkmal Schicht auf: Falls ein Bonus für Abkunft

aus niedriger Schicht gegeben würde, würde nicht eine durchschnittliche Verschlechterung der Zwischen-Examina zu erwarten sein; in Medizin wäre sogar eher der entgegengesetzte Effekt zu erwarten. Diese Feststellung gilt allerdings nur für die hier untersuchte Wirklichkeit; Änderungen des Bildungsverhaltens auch infolge von Auslesemaßnahmen können solche Feststellungen u.U. schnell überholen.

Der bei der Vorhersage der Examenszensuren durch die Abiturnoten auftretende paradoxe Effekt, daß die dem Studienfach korrespondierenden Schulfächer nur relativ geringen prognostischen Wert besitzen, wurde auf Ausleseeffekte durch die Studienfachwahl zurückgeführt. Im folgenden Kapitel (4) soll u.a. die Analyse dieser Ausleseeffekte erfolgen.

4. Weitere Untersuchungen im Rahmen des linearen Modells

4.1 Prüfung des Kriteriums "Zensur im Zwischenexamen" auf Gleichabständigkeit der Skalenstufen und Dimensionalität

4.1.1 Rationale

Eine direkte Prüfung der meßtheoretischen Eigenschaften der Kriteriumskala, d.h. der Repräsentation der Relationen innerhalb des wahren Merkmals durch die hier verwendeten Meßwerte ist nicht möglich. Jedoch läßt sich die Hypothese untersuchen, ob sich einerseits die Beziehungen zwischen den Prädiktoren und dem Kriterium im Modell der Regressionsanalyse, d.h. unter Annahme eines eindimensionalen, intervallskalierten Kriteriums in optimaler Weise darstellen lassen, oder ob andererseits bessere Anpassung des Modells an die Daten möglich ist, wenn diese Voraussetzungen aufgegeben werden.

Die empirische Prüfung dieser Hypothese läßt sich unter Einsatz einer Diskriminanzanalyse durchführen. Dazu werden die Stufen des - bei der Regressionsanalyse quantitativen - Kriteriums als Gruppenmerkmale aufgefaßt; das Kriterium wird also hierbei ein qualitatives Merkmal, bei dem keine Annahmen über die Gleichabständigkeit der Skalenstufen oder die Dimensionalität bestehen. Nahezu deckungsgleich mit dieser Methode wäre die Auflösung des Kriteriums in Dummy-Variablen und die Berechnung von kanonischen Korrelationen zwischen Prädiktoren und aus dem Kriterium gebildeten Dummyvariablen, d.h. die Durchführung einer "kanonischen Diskriminanzanalyse" (vgl. HOPE, 1975, VAN DE GEER, 1971)[1].

Der Versuchsaufbau wird in Abb. 4.1 veranschaulicht:

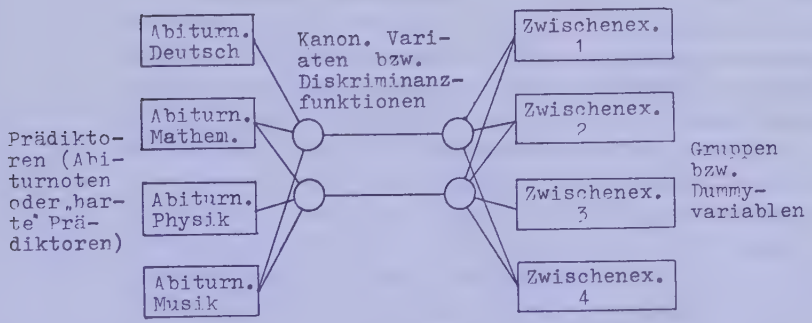


Abb. 4.1: Die Vorhersage der Zensur des Zwischenexamens mit einer Diskriminanzanalyse bzw. kanonischen Diskriminanzanalyse

1) Das Vorgehen ist jenem Verfahren bei der Regressionsanalyse analog, bei dem ein eigentlich quantitativer Prädiktor in Dummyvariablen aufgelöst wird um prüfbar zu machen, ob durch die Dummyvariablen mehr Varianz erklärt wird als durch den quantitativen Prädiktor (vgl. KERLINGER und PEDHAZUR, 1973).

Die Fragen, die an die Ergebnisse einer Diskriminanz - oder kanonischen Diskriminanzanalyse gestellt werden, sind folgende:

1. Wieviele signifikante Trennfunktionen gibt es; liegen die Gruppen im Diskriminanzraum auf einer Geraden, d.h. lassen sie sich auf einer Dimension abbilden, oder läßt sich die räumliche Anordnung der Gruppen nicht eindimensional beschreiben?

2. Haben die Gruppenzentroide auf der Diskriminanzfunktion - falls es nur eine gibt - gleiche Abstände voneinander, d.h. ist die Beziehung zwischen der Diskriminanzfunktion und dem - jetzt wieder - quantitativ aufgefaßten Gruppenkriterium linear? Eine solche lineare Beziehung soll in Abb. 4.2 veranschaulicht werden.

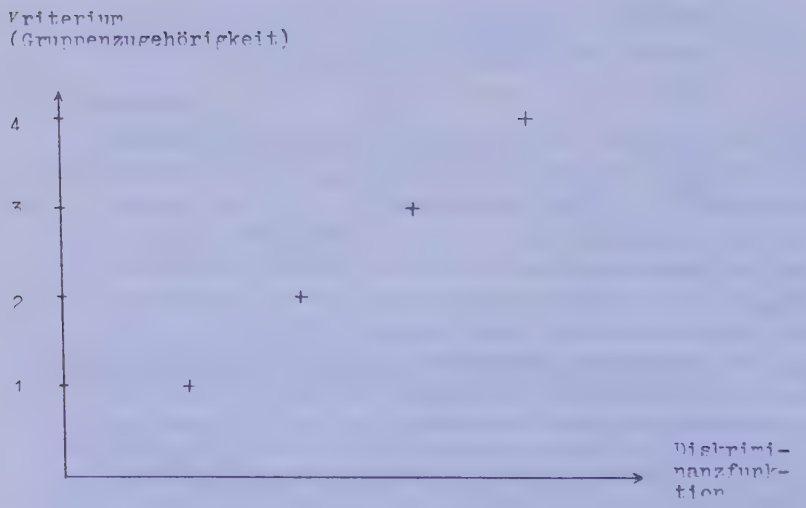


Abb. 4.2: Lineare Beziehung zwischen Gruppenzentroiden auf der Diskriminanzfunktion und dem quantitativ aufgefaßten Gruppenkriterium

Es ist einleuchtend, daß bei Vorliegen von einer Diskriminanzfunktion und linearer Beziehung zwischen Funktion und quantitativem Kriterium wie z.B. Abb. 4.2 eine Regressionsanalyse eben so gut zur Vorhersage des Kriteriums geeignet ist wie eine Diskriminanzanalyse.[1] In diesem Fall wäre die Regressionsanalyse als das sparsamere und durch die Schätzung von weniger Parametern effizientere Modell der Diskriminanzanalyse vorzuziehen.

4.1.2 Ergebnisse und Diskussion

Die sieben größten Studienfächer wurden jeweils zwei Diskriminanzanalysen unterzogen; eine mit den harten Prädiktoren und eine mit den Abiturnoten als Prädiktoren. Die Gruppen waren jeweils durch die Stufen des Kriteriums "Zensur im Zwischenexamen" definiert. Es wurde das Programm "Discriminant" des Pakets SPSS (NIE et all. 1970) benutzt. Die Prädiktoren wurden ähnlich wie bei einer multiplen Regressionsanalyse schrittweise nach ihrer Fähigkeit zur Reduzierung der Residualvarianz in die Analyse eingeführt. In einzelnen Fällen wurden sehr schwach besetzte Klassen des Kriteriums der nächsten Klasse zum Mittelwert zugeschlagen. Überhaupt sind die Zensurengruppen unterschiedlich stark besetzt; schwach vertreten sind häufig die Zensuren 1 und 4. Da diese Tatsachen bei der Interpretation berücksichtigt werden müssen, werden die Umfänge der Gruppen bei den Ergebnissen mit angegeben. Die Berücksichtigung fehlender Angaben mußte durch fallweisen Ausschluß aller Personen mit mindestens einer fehlenden Angabe erfolgen, dadurch wurden die Fallzahlen wesentlich reduziert. Weil dadurch bei den kleineren Studienfächern die kleinsten, vernünftigen Stichprobenumfänge unterschritten sind, werden die Ergebnisse der drei größten Fächer ausführlich, die der restlichen Fächer nur summarisch wiedergegeben.

---

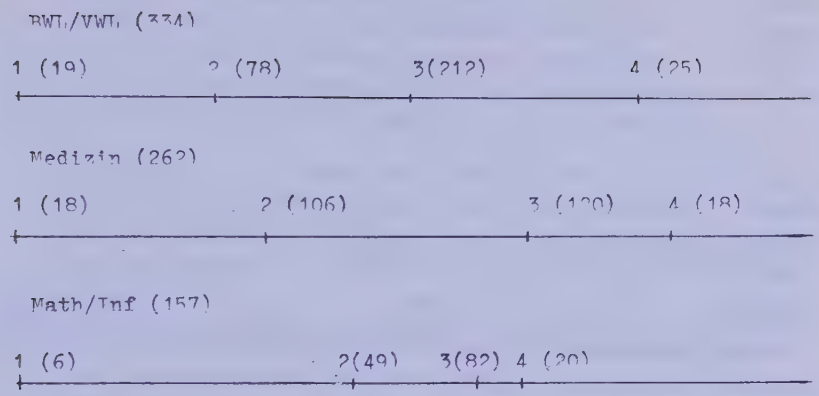
1) Wiederum im analogen Fall eines in Dummyvariablen aufgelösten quantitativen Prädiktors bei einer Regressionsanalyse (vgl. KERLINGER et al. 1973) würde das heißen, der Koeffizient Eta-Quadrat ist nicht größer als r-Quadrat.

Zunächst die Befunde zur ersten der gestellten Fragen: Bei den Abiturnoten als Prädiktoren hat jedes Studienfach eine, bei den "harten" Prädiktoren hat jedes Studienfach höchstens eine - Deutsch und Physik keine - mindestens auf dem 0,05 Niveau signifikante Trennfunktion. Die der ersten Diskriminanzfunktion zugehörigen Eigenwerte umfassen stets den größten Teil der Spur (um 70 %), während der zweite Eigenwert bereits stark abfällt und die Reihe der einen linearen Abfall zeigenden Eigenwerte einleitet. Somit wird die Wahl von höchstens einer Diskriminanzfunktion außer vom statistischen, auch durch ein praktisches Kriterium gestützt; man kann also von der Darstellbarkeit der Gruppen im Diskriminanzraum auf einer Dimension ausgehen.

Zur Beantwortung der zweiten Frage werden die Gruppenzentroide auf den jeweils ersten Diskriminanzfunktionen in Abb. 4.3 dargestellt. Die Gruppe 1 besteht aus den Studierenden mit der Zensur im Zwischenexamen 1, usw. In Klammern sind die Stichprobenumfänge der Studienfächer, sowie die Anzahlen in den betr. Gruppen angegeben. In Abb. 4.4 ist ein Fall der guten bis durchschnittlichen Annäherung an Gleichabständigkeit (BWL/VWL mit "harten" Prädiktoren), in Abb. 4.5 der Fall der schlechtesten Anpassung an Gleichabständigkeit (Math/Inf mit Abiturnoten als Prädiktoren) im Zusammenhang der Trennfunktion mit dem quantitativen Gruppenkriterium graphisch dargestellt, weil in dieser Form eine Beurteilung am leichtesten möglich ist.

Die Annäherung an Gleichabständigkeit der Gruppen bzw. Stufen des Kriteriums erscheint zufriedenstellend, vor allem wenn man berücksichtigt, daß sich die stärksten Abweichungen bei den Gruppen mit dem kleinsten Umfang befinden. Mindestens sind keinerlei Umkehrungen in der Rangfolge zu beobachten. Eine gewisse Tendenz, daß die Gruppen 3 und 4 eher dichter, und die Gruppen 1 und 2 eher weiter auseinanderliegen, deutet sich allerdings an. Dennoch dürften die Abweichungen nicht schwerwiegend genug sein, um die Angemessenheit einer multiplen Regressionsanalyse zu verwerfen und die größere statistische Unsicherheit durch das komplexere Modell der Diskriminanzanalyse in Kauf zu nehmen.

Abiturnoten



"Harte" Prädiktoren

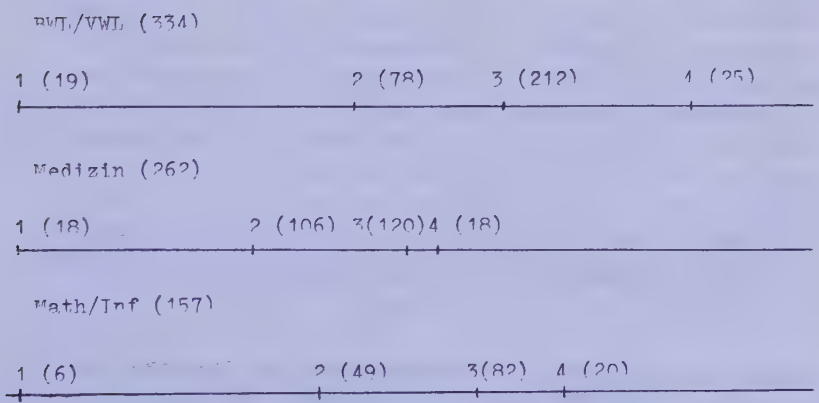


Abb. 4.3: Lage der Gruppenzentroide auf der jeweiligen ersten Diskriminanzfunktion, in Klammern die Gruppenumfänge

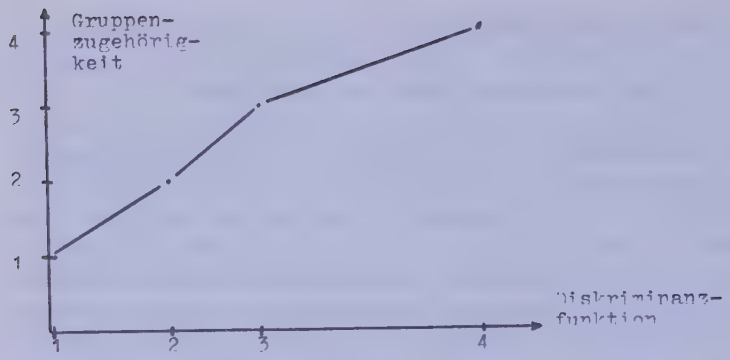


Abb. 4.4: Beispiel eines Zusammenhangs von Diskriminanzfunktion und quantitativ aufgefaßtem Gruppenkriterium mit guter Annäherung an Linearität (BWL/VWL, "harte" Prädiktoren)

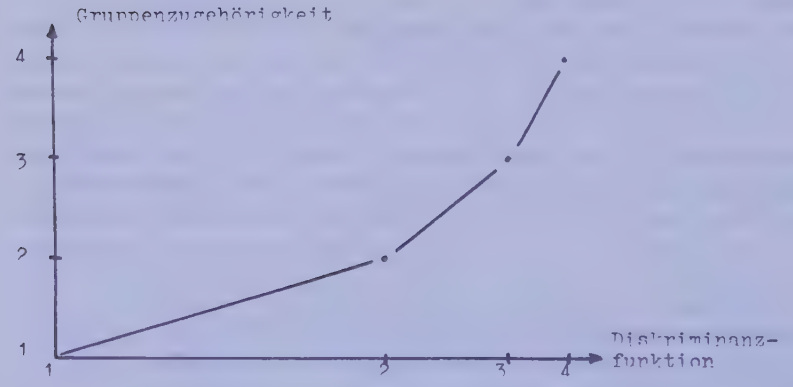


Abb. 4.5: Beispiel eines Zusammenhangs von Diskriminanzfunktion und quantitativ aufgefaßtem Gruppenkriterium mit der am wenigsten guten Annäherung an Linearität (Math/Inf, Abiturnoten)

Für die restlichen, bisher nicht beschriebenen Studienfächer ist das Bild ähnlich. Teilweise sind die Abweichungen etwas größer und es treten sogar einzelne Inversionen der Rangfolge bei besonders schwach besetzten Gruppen auf, jedoch dürften diese Abweichungen aufgrund der geringen Stichprobenumfänge bereits nach dem Zufall zu erwarten sein.

Insgesamt ist festzustellen, daß die hier durchgeführten Analysen keine Hindernisse aufdeckten, das Kriterium "Zensur im Zwischenexamen" als eindimensionale Intervallskala aufzufassen und im Modell der multiplen Regression zu analysieren.

4.2 Unterscheidung der Studienfächer aufgrund der Abiturnoten

Bei der Darstellung der Unterschiede zwischen den Studienfächern bezüglich iherer Mittelwerte und Streuungen in den Prädiktoren in Kap. 3.4.1 (Tab. 3.5) tauchte die Überlegung auf, die Abiturnoten als Prädiktoren des gewählten Studienfachs zu benutzen. Ziel dieses Vorgehens ist es zum einen, weitere Erkenntnisse über die prognostische Validität der Abiturnoten zu gewinnen, die man offensichtlich nicht nur in bezug auf die Vorhersage des Studienerfolgs, sondern auch in bezug auf die Vorhersage der Studienfachwahl sehen muß. Und beide Zusammenhänge sind sicher insofern nicht unabhängig voneinander, als der Teil der Varianz der Abiturnoten, der das Fach vorhersagt, nicht mehr zur Vorhersage des Studienerfolgs zur Verfügung steht (vgl. 4.3). Zum anderen können Ergebnisse erwartet werden, die eventuell eine Anwendung bei der Beratung von Abiturienten erlauben. Dabei wären dann zwei Gesichtspunkte zu beachten.

1. Die Abiturnoten lassen die Wahl eines bestimmten Studienfachs günstig oder nicht günstig erscheinen, insoweit sie größere oder geringere Ähnlichkeit mit den Abiturnoten derjenigen Studenten aufweisen, die dieses Fach gewählt und bereits ein Zwischenexamen abgelegt haben.[1)]

---

1) Allerdings kommt bei einzelnen Studienfächern durch den n.c. "unechte", für die Beratung eher irrelevante Varianz zum Tragen.

2. Des weiteren lassen die Abiturnoten in einem bestimmten Fach eher ein besseres oder schlechteres Zwischenexamen - entsprechend den Ergebnissen von Kap. 3.4.1 - erwarten.

Die Hypothese, daß das gewählte Studienfach durch die Abiturnoten vorhersagbar ist, soll nachfolgend mit Hilfe von Diskriminanz- bzw. kanonischen Diskriminanzanalysen sowie von Regressionsanalysen untersucht werden. Es wurden die Programme "Discriminant", "Cancorr" und "Regression" des Pakets SPSS (NIE et al. 1970) benutzt. Im Gegensatz zu der von ORLIK (1967) in der gleichen Situation benutzten Profilmethode, die auch die Erfassung von konfigurationalen "physiognomischen" Effekten - sofern sie vorhanden sind - erlaubt, besteht hier die Einschränkung darauf, daß die Studienfächer aufgrund von linearen Kombinationen der Abiturnoten unterschieden werden. Andererseits werden hier - im Vorteil gegenüber der von ORLIK verwendeten Methode - auch die Kovarianzen der Prädiktoren berücksichtigt und verarbeitet.

4.2.1 Ergebnisse

Bei der Diskriminanzanalyse der 1613 Studierenden der 7 größten Studienfächer (hier wiederum eine Verringerung der Stichprobe durch fallweisen Ausschluß fehlender Daten) mit den Studienfächern als zu unterscheidenden Gruppen ergaben sich folgende Befunde:

- Von den Abiturnoten trugen die Fächer Erdkunde und Musik am wenigsten zur Differenzierung der Studienfächer bei. Dies ergab sich bei schrittweiser Einführung der Prädiktoren sowohl nach dem Kriterium von WILKS als auch dem Kriterium der Minimierung der Residualvarianz.

- Fünf der berechneten Diskriminanzfunktionen waren auf dem 0,05 Niveau signifikant; zwei dieser Funktionen wurden nach dem von COOLEY und LOHNES (1971) vorgeschlagenen Kriterium (kanonische Korrelation größer als 0,30) sowie nach einem dem Scree-Test bei der Faktorenanalyse

(CATELL, 1966) analogen Test als relevant erachtet.

- Die den zwei relevanten Dimensionen zugehörigen Eigenwerte umfassen zusammen 85 % der Spur, die entsprechenden kanonischen Korrelationen sind 0,54 und 0,41.

- Die (Re-)Klassifizierung der Fälle war nach den zwei relevanten Dimensionen in 42 %, nach den fünf signifikanten Dimensionen in 44 % der Fälle - das ist jeweils signifikant bei 0,0001 Zufallswahrscheinlichkeit - zutreffend. Dabei war die a priori Wahrscheinlichkeit der Zugehörigkeit zu den Studienfächern nicht spezifiziert, d.h. als gleich für alle Fächer angenommen.

- Die Ladungen der Abiturnoten und Studienfächer auf den kanonischen Variaten sind in Tabelle 4.1 angegeben:

Tabelle 4.1: Die Ladungen der jeweils ersten beiden kanonischen Variaten

Koeffizienten der Abiturnoten

	Variate 1	Variate 2
Deutsch	-.35	.17
Englisch	-.28	.55
Erdkunde	-.02	-.02
Mathematik	.77	-.15
Physik	.36	.34
Biologie	.04	.30
Musik	-.09	.21

Koeffizienten der Studienfächer

	Variate 1	Variate 2
BWL/VWL	.83	.69
Medizin	.65	-.38
Math/Inf	-.09	0
Deutsch	.95	-.06
Französ.	.77	-.20
Chemie	.16	0
Wi.-Päd.	.38	.36

- Zur Interpretation der den Prädiktoren (Abiturnoten) und Kriterien[1] (Studienfächern) gemeinsamen Varianz genügt es nicht, Bezug auf die kanonischen Korrelationen zu nehmen, da diese nur die zwischen den jeweiligen kanonischen Variaten, nicht aber die zwischen den beiden Variablengruppen gemeinsame Varianz angeben. Deshalb wurden in Tab. 4.2 die Korrelationen der Kriterien mit den kanonischen Variaten der Prädiktoren, die sog. Redundanzen nach STEWART und LOVE (1968, nach COOLEY und LOHNES, 1971) berechnet.[2] Ihre Quadrate geben an, wieviel die kanonischen Variaten zur Differenzierung der einzelnen Studienfächer beitragen.

  Die für die kanonischen Variaten aufsummierten Quadrate dieser einzelnen Redundanzen ergeben die Redundanzen zwischen jeweils einer kanonischen Variate der Prädiktoren und der Gesamtheit der Kriterien. Diese Gesamtredundanzen und die ihnen entsprechenden Anteile an der gesamten Kriterienvarianz sind in den beiden letzten Zeilen der Tab. 4.2 enthalten. Danach erklären die erste kanonische Variate 4,8 % und die zweite kanonische Variate der Prädiktoren 4,2 %, beide zusammen also 9 % der gesamten Kriterienvarianz.

- In Abb. 4.6 ist die Lage der Kriterien und der Prädiktoren auf den zugehörigen kanonischen Variaten graphisch dargestellt, nachdem die Achsen gleichzeitig für Prädiktoren und Kriterien graphisch rotiert worden waren. Zu beachten ist, daß alle Angaben der Tabellen 4.1 und 4.2 sich auf die unrotierten Variaten beziehen, und daß lediglich der durch beide Variaten gemeinsam erklärte Anteil von 9 % der Kriterienvarianz durch die

---

1) Es wird hier aus einleuchtenden Gründen durchgängig von Prädiktoren und Kriterien gesprochen, obwohl das Modell völlig symmetrisch ist und lediglich 1. und 2. Variablengruppe unterscheidet.

2) Nur die Redundanzen in dieser Richtung - obwohl vom Modell her wiederum auch die Redundanzen in der anderen Richtung gleichberechtigt sind - sollen in diesem Abschnitt berücksichtigt werden.

Tabelle 4.2: Die Redundanzen der Kriterien mit den kanonischen Variaten der Prädiktoren

	Variate 1	Variate 2
BWL/VWL	0,08	0,34
Medizin	0,02	- 0,31
Math/Inf	- 0,38	- 0,05
Deutsch	0,29	- 0,07
Französisch	0,21	- 0,12
Chemie	- 0,10	- 0,03
Wirtschaftspädagogik	0,02	0,13
gesamte Redundanz der Variate mit allen Kriterien	0,29	0,25
erklärter Anteil der Kriterienvarianz	4,8 %	4,2 %

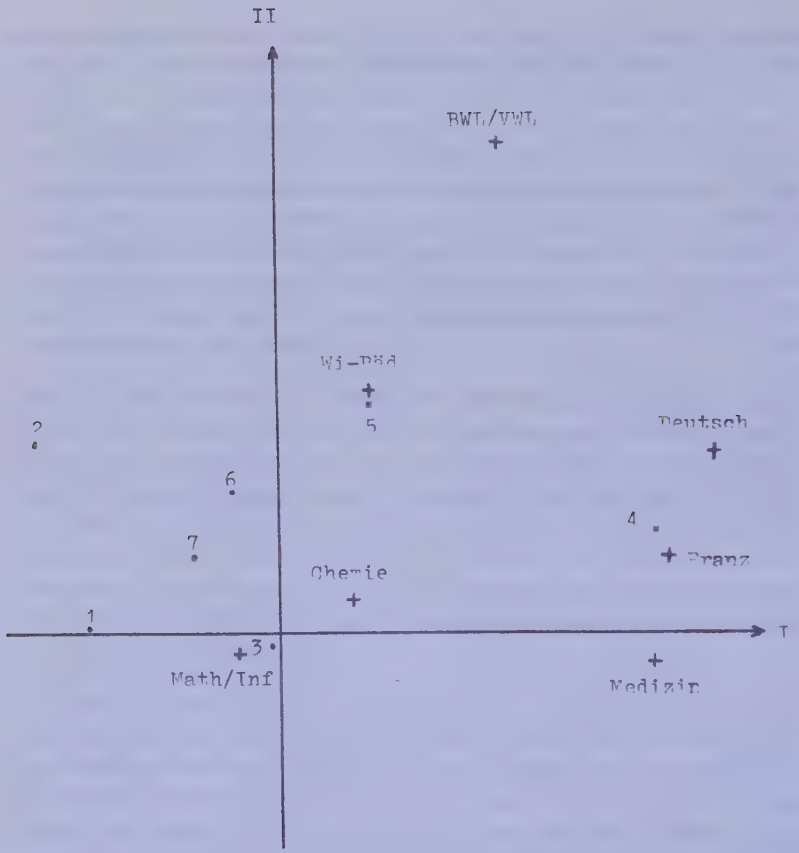


Abb. 4.6: Lage der Studienfächer (Symbol +) und der Abiturnoten (Symbol ·, Erklärung der Ziffern s.u.) auf den zugehörigen, gemeinsam rotierten, kanonischen Variaten

Abiturnoten: 1=Deutsch, 2=Englisch/1. Fremdspr., 3=Erdkunde, 4=Mathematik, 5=Physik, 6=Biologie, 7=Musik

Rotation unberührt bleibt. Die Abb. 4.6 soll als Basis für die Interpretation der Lage der Studienfächer im Diskriminanzraum dienen. (Zu beachten ist, daß wegen der Polung der Zensurenskala negative Ladungen der Abiturnoten guten Zensuren entsprechen).

Zur Ergänzung der vorstehend berichteten Diskriminanzanalyse, die das angemessene statistische Modell zur Prüfung der eingangs von 4.2 genannten Hypothese darstellt, wurden Regressionsanalysen berechnet, die Teilaspekte der Diskriminanzanalyse behandeln. Dabei wird je ein einzelnes Studienfach den restlichen Fächern gegenübergestellt. Dieser Weg wurde gewählt, um sätzliche beschreibende Informationen zu erhalten und die hier gefundenen Ergebnisse in leichterer Weise parallel zu den Befunden des Kap. 3.4.1, bei denen ja auch jeweils ein einzelnes Studienfach untersucht wurde, darstellen zu können. Technisch gesehen wurden also Regressionsanalysen mit den Abiturnoten als Prädiktoren und einem dichotomen Merkmal als Kriterium gerechnet, das die Information trug, ob der betreffende Studierende dem untersuchten Fach angehörte oder nicht.[1)]

Die Frage, ob in Regressionsanalysen ein dichotomes Kriterium verwendet werden sollte, ist umstritten; neuere Untersuchungen (nach MEREDITH et al., 1974, S. 461 und 487) scheinen eher einer positiven Antwort zuzuneigen. Unabhängig von der Antwort auf diese Frage sind jedoch weitere Modellvoraussetzungen der Regressionsanalyse (für eine Diskriminanzanalyse würde das entsprechende gelten) vor allem dadurch nicht erfüllt, daß die Gruppe der nicht dem untersuchten Studienfach angehörenden Studierenden durch die verschiedenen Studienfächer innere Abhängigkeiten und damit korrelierte Fehler enthalten dürfte. Insofern müssen die durchgeführten Signifikanztests als zu große Effizienz vorspiegelnd und deshalb als rein beschreibend gewertet werden.

Die Ergebnisse nach dem bereits in Kap. 3 benutzten und beschriebenen Verfahren der schrittweisen Repressionsanalyse sind in Tab. 4.3 dargestellt. Jeder zusätzlich in die Regressionsgleichung eingeführte Prädiktor erklärt mindestens 0,5 % weitere Kriteriumsvarianz,

---

1) Eine Regressionsanalyse mit dichotomem Kriterium ist nach VAN DE GEER (1971) und HOPE (1975) einer Diskriminanzanalyse mit zwei Gruppen exakt analog, die hier also ebenso gut hätte stehen können.

Tabelle 4.3: Die Vorhersage der einzelnen Studienfächer innerhalb der Studierenden von 8 Fächern (N = 2036) durch die Abiturnoten

Kriterium	Prädiktoren	Validität	Multiple Korrelation	Prozent erklärte Varianz
BWL/VWL vs. Restfächer	Physik (-) [1]	0,26	0,26	6,9
	Englisch(-)	0,26	0,33	10,8
	Biologie (-)	0,22	0,34	11,7
	Deutsch (-)	0,23	0,35	12,1
Medizin vs. Restfächer	Biologie	- 0,25	0,25	6,1
	Englisch	- 0,16	0,27	7
	Musik	- 0,15	0,28	7,7
Mathematik/Inf. vs.Restfächer	Mathematik	- 0,37	0,37	13,7
	Deutsch[2]	0,03	0,39	15,1
Deutsch vs. Restfächer	Deutsch	- 0,21	0,21	4,6
	Mathematik(-)[3]	0,19	0,33	10,9
Französisch vs. Restfächer	Engl./1.Fremdspr.	- 0,24	0,24	5,6
	Mathematik(-)[3]	0,12	0,29	8,6
Chemie vs. Restfächer	Biologie	- 0,09	0,09	0,9
	Musik (-)	0,04	0,11	1,4
Physik vs. Restfächer	Physik	- 0,18	0,18	3,2
	Musik (-)	0,05	0,20	4,1
Wirtschafts-pädagogik vs. Restfächer	Englisch	0,11	0,11	1,1
	Physik	0,09	0,13	1,6

1) Bei Abiturnoten mit nachgestelltem Minuszeichen sind schlechte Zensuren mit der Wahl des betreffenden Studienfachs verbunden.

2) Suppressorwirkung nach herkömmlicher Definition

3) (Partielle) Suppressorwirkung nach der Definition von CONGER (1974)

einen Zuwachs, der mindestens auf dem 0,01 Niveau "signifikant" ist. Alle Prädiktoren mit höherer Validität wurden auch für die Regressionsgleichung ausgewählt, sind also in der Tabelle enthalten. Nicht aufgeführte Abiturnoten korrelieren also zu Null oder minimal mit dem Kriterium.

4.2.2 Diskussion

Die Hypothese, daß sich das gewählte Studienfach aus der Kenntnis der Abiturnoten vorhersagen läßt, wird durch die Ergebnisse der Diskriminanzanalyse bestätigt. Es traten fünf signifikante Diskriminanzfunktionen auf, und die (Re-) Klassifizierung der Fälle ist signifikant besser als der Zufall. Zwei der Diskriminanzfunktionen wurden als relevant erachtet. Berechnet man nach der Methode von STEWART und LOVE (1968, nach COOLEY und LOHNES, 1971) die Redundanzen zwischen den Kriterien und den zwei relevanten kanonischen Variaten der Prädiktoren, so zeigt sich, daß insgesamt 9 % der Varianz der Kriterien durch diese Variaten erklärt wird. Dieser Prozentsatz läßt sich analog dem Quadrat einer multiplen Korrelation interpretieren, entspricht also einer multiplen Korrelation von 0,30. Damit bewegt sich die Vorhersage des Studienfachs etwa in der gleichen Größenordnung wie die Vorhersage des Studienerfolgs, (vgl. 3.4.1) wobei sich allerdings die Prozentsätze an erklärter Varianz einmal auf ein einzelnes und einmal auf eine Gruppe von Kriterien beziehen. (Die Erklärung von Varianz der Gruppe der Kriterien kann man sich in 2 Schritten vorstellen: zunächst wird ein Teil der Varianz in Form der kanonischen Variaten der Kriterien extrahiert und danach wird die Übertragung von Varianz der kanonischen Variaten der Prädiktoren auf die der Kriterien bestimmt.)

Die Redundanzen der beiden Variaten der Prädiktoren mit den einzelnen Kriterien in Tab. 4.2 zeigen, daß die kleineren Studienfächer im zweidimensionalen System nicht so gut dargestellt werden können wie die großen. Eine bessere Separierung der kleineren Studienfächer leisteten die weiteren drei signifikanten Dimensionen; jedoch wurde das Ausmaß dieser Verbesserung für insgesamt nicht sehr wesentlich erachtet.

Als erste unrotierte Trennfunktion ergibt sich eine bipolare, durch bessere Mathematikzensuren gegenüber schlechteren Zensuren der sprachlichen Fächer gekennzeichnete, Dimension, die also einen differentiellen Aspekt der Schulnoten zum Inhalt hat. Sie trennt vor allem das Studienfach Math/Inf von den sprachlichen Studienfächern. Die zweite unrotierte Trennfunktion stellt eine unipolare, von fast allen Abiturnoten geladene, Dimension dar, die einen generellen Aspekt der Abiturnoten enthält. Sie trennt vor allem BWL/VWL mit weitem Abstand von den anderen Studienfächern; am weitesten entfernt liegt Medizin. D.h. das Studienfach BWL/VWL ist durch generell schlechtere Zensuren gegenüber den anderen Studienfächern gekennzeichnet.

Die geringfügige Rotation der Achsen in Abb. 4.6 um 25° nach rechts veränderte das Bild nicht sehr wesentlich, wenn auch die Interpretation etwas erleichtert wurde. Plausibel wird die Geringfügigkeit der notwendigen Rotation durch die bereits recht gute Interpretierbarkeit der unrotierten Lösung, sowie vielleicht durch die weniger steile Abstufung in der Varianzstärke der extrahierten Dimensionen z.B. gegenüber den meisten Faktorenanalysen. Die Abbildung 4.6 bestätigt eine differentielle Abiturnotendimension - gekennzeichnet durch bessere Mathematikzensuren gegenüber den sprachlichen Fächern -, die das Studienfach Math/Inf von Deutsch, Französisch und Medizin trennt. Die zweite, eine generelle Abiturnotendimension - geladen vor allem von Physik, Englisch und Biologie - trennt BWL/VWL mit weitem Abstand von den anderen Studienfächern. Am weitesten zum

entgegengesetzten Pol liegen Medizin und Math/Inf, die auf dieser Dimension durch die besten Noten gekennzeichnet sind.

Auffällig ist die Nähe des Studienfachs Medizin zu den sprachlichen Fächern, unterschieden von ihnen nur die etwas besseren Zensuren auf der generellen Notendimension.

Wenig prägnant im Mittelfeld liegen die Studienfächer Wirtschaftspädagogik und Chemie, in der Tendenz jeweils nahe den verwandten Fächern BWL/VWL bzw. Math/Inf. Dieser Befund der wenig prägnanten Kennzeichnung dieser Fächer wird durch die anschließend besprochenen Ergebnisse der Regressionsanalysen (Tab. 4.3) bestätigt.

Die Regressionsanalysen waren nicht als statistische Verfahren zur Beantwortung der grundsätzlichen Frage, ob sich die Studienfächer aus den Abiturnoten vorhersagen lassen, eingesetzt worden. Vielmehr sollten sie weitere beschreibende Informationen, die besser auf die Situation der einzelnen Fächer abgestimmt sind, liefern. Die einzelnen Regressionsanalysen sind dadurch, daß jeweils der im wesentlichen gleiche Datensatz analysiert wird, nicht unabhängig voneinander.

Die Ergebnisse der Tab. 4.3 zeigen u.a. insofern Übereinstimmung mit Abb. 4.6, als auch hier Math/Inf, Deutsch und Medizin die am besten vorhersagbaren Fächer sind.

Nachstehend folgen einige Bemerkungen zu den einzelnen Studienfächern:

- BWL/VWL wird durch generell schlechte Noten, vor allem in Physik, Englisch und Biologie geprägt.
- Medizin zeichnet sich durch generell gute Noten, vor allem in Biologie, Englisch und Musik aus.
- Die Studienfächer Math/Inf, Deutsch und Französisch zeigen jeweils analoge Muster. Sie lassen sich durch gute Noten im korrespondierenden Abiturfach (Mathematik bzw. Deutsch bzw. Englisch/1. Fremdsprache) im Verhältnis zu schlechteren Noten im "entgegengesetzten" Fach (Deutsch bzw. Mathematik) kenn-

zeichnen. Dabei ist die Wirkung des jeweils "entgegengesetzten" Schulfachs auf einen Suppressoreffekt im konventionellen Sinn (HORST, 1971) bei Math/Inf oder auf einen (partiellen) Suppressoreffekt nach der neueren Definition von CONGER und JACKSON (1972, 1974) bei Deutsch und Französisch zurückzuführen. D.h. die Wirkung des "entgegengesetzten" Abiturfachs beruht jeweils voll bzw. teilweise auf der Unterdrückung von irrelevanter Varianz im homologen Abiturfach. Z.B. unterdrückt beim Studienfach Math/Inf der Suppressor "Deutschnote im Abitur" vermutlich die in der Mathematikzensur enthaltene, für die Vorhersage des Studienfachs Math/Inf irrelevante, sprachliche Komponente.

- Die Fächer Chemie, Wirtschaftspädagogik und Physik (dieses Fach war wegen zu geringen Stichprobenumfangs nicht in die Diskriminanzanalyse einbezogen) erwiesen sich als nicht oder kaum vorhersagbar.

Die Höhe der Vorhersage liegt bei den größeren fünf Fächern in ähnlicher Höhe wie bei der Vorhersage der Examenszensuren (0,28 - R - 0,39). Die in den zugehörigen Grundgesamtheiten zu erwartende Schrumpfung wurde nicht berechnet, denn sie wäre wegen des jeweils gleichen Stichprobenumfangs bei allen Fächern in gleicher Weise wirksam. Außerdem wäre die nach der angemessenen Schätzformel berechnete Schrumpfung wegen der umfangreichen Stichprobe nur sehr gering, während hingegen die wirklich zu erwartende Schrumpfung wegen der Abweichung von den Modellannahmen wiederum in nicht zu spezifizierender Weise höher als die nach der Schätzformel berechnete ausfallen würde.

4.3 Gleichzeitige Vorhersage von Studienfach und Studienerfolg

Dieser Ansatz wurde gewählt, um die Anteile der Abiturnoten an der Vorhersage des Studienerfolgs (einzeln untersucht in 3.4.1) und des Studienfachs (einzeln untersucht in 4.2) in einer gemeinsamen Analyse gegenüberzustellen. Zusätzlich sollte geprüft werden, ob die diskriminanzanalytische Unterscheidung gleichzeitig der Zensuren in einem Fach der Zugehörigkeit zu diesem oder einem anderen Fach (s.u.) zu einer eindimensionalen oder mehrdimensionalen räumlichen Trennung (ähnlich wie in 4.1) führen würde, und welche Konfiguration der Gruppen sich gegebenenfalls zeigen würde. Es wurden Diskriminanz- und kanonische Diskriminanzanalysen mit den bereits genannten Programmen nach folgendem Muster gerechnet: Der erste Variablensatz der Prädiktoren besteht aus den Abiturnoten. Der zweite Variablensatz der Kriterien bzw. der Gruppen besteht aus den Zensuren im Zwischenexamen eines bestimmten Fachs und der Zugehörigkeit zu diesem Fach oder nicht, wie z.B. nachfolgend in Abb. 4.7 für das Studienfach BWL/VWL gegenüber den Restfächern dargestellt:

BWL Vordiplom = 1	BWL VD = 2	BWL VD = 3	BWL VD = 4	nicht BWL studierend

Abb. 4.7: In der Diskriminanzanalyse zu unterscheidende Gruppen bzw. Merkmale, aus denen die Dummyvariablen für die Kriterien der kanonischen Diskriminanzanalyse gebildet werden.

4.3.1 Ergebnisse

Da auch hier die Stichproben durch fallweisen Ausschluß der fehlenden Angaben stark reduziert waren, werden nur die Studienfächer BWL/VWL, Medizin, Math/Inf, Deutsch und Französisch berücksichtigt. Bei den restlichen Fächern war ohnehin eine Vorhersage des Studienfachs durch die Abiturnoten kaum möglich. Zu beachten ist wiederum der stark unterschiedliche Umfang der in den Diskriminanzanalysen zu unterscheidenden Gruppen; es fällt jeweils die sehr umfangreiche Restgruppe der nicht in diesem Fach studierenden Personen heraus. Auch hier besteht wie bei den Regressionsanalysen von 4.2 eine gewisse Abhängigkeit der einzelnen Analysen durch Überlappung der jeweils analysierten Datensätze.

Für jedes Fach ergaben sich zwei signifikante Dimensionen, die mit den Gruppenzentroiden nach graphischer Rotation in den Abbildungen 4.8 bis 4.12 dargestellt wurden. Die Rotationen erfolgten auf das Ziel hin, daß eine Dimension möglichst rein zwischen den Zensuren und die andere möglichst rein die Fachzugehörigkeit unterschied. Die Rotationen waren unnötig für Französisch, minimal für Math/Inf und Deutsch und wesentlich (ca. 45°) für Medizin und BWL/VWL.

Es wurden die Varianzanteile berechnet und in Tabelle 4.4 zusammengefaßt, die die rotierten Diskriminanzfunktionen aus der gesamten Varianz der Prädiktoren extrahieren.

4.3.2 Diskussion

Eine Dimension allein ist nicht ausreichend zur jeweiligen Trennung aller Gruppen. Die Restgruppe der nicht in einem Fach Studierenden läßt sich nicht mit der gleichen Dimension, die die Zensurengruppen des betreffenden Fachs voneinander trennt, von den Studierenden dieses Fachs unterscheiden.

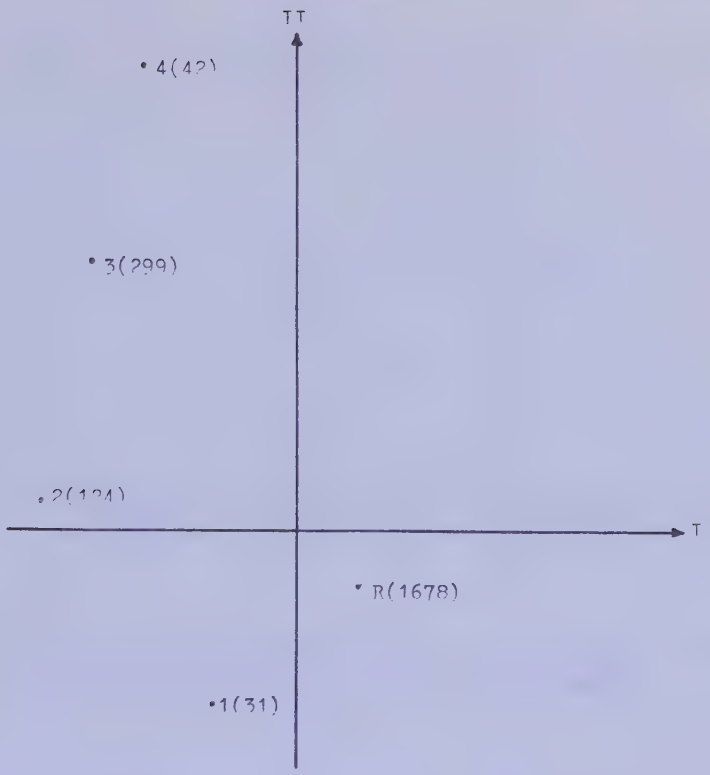


Abb. 4.8: Trennung zwischen Studierenden der BWL/VWL mit den Vordiplomzensuren 1,2,3 und 4 und den Studierenden der anderen Fächer (R) durch die Abiturnoten

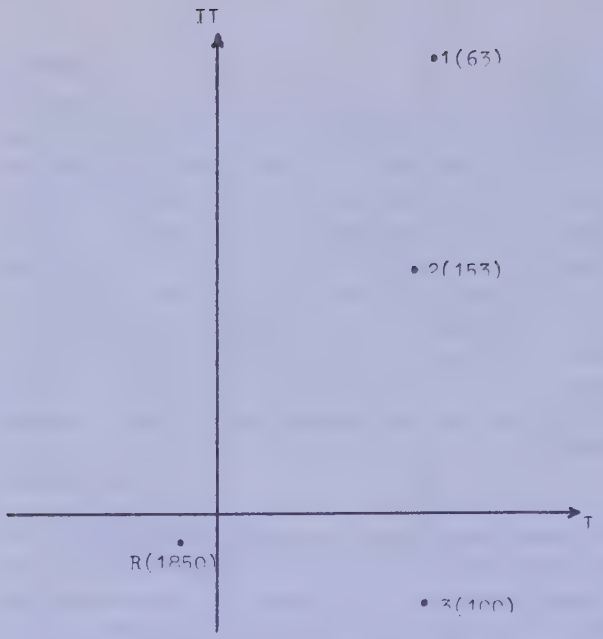


Abb. 4.9: Trennung zwischen Studierenden der Medizin mit den Physikumzensuren 1,2 und 3 und den Studierenden der anderen Fächer (R) durch die Abiturnoten

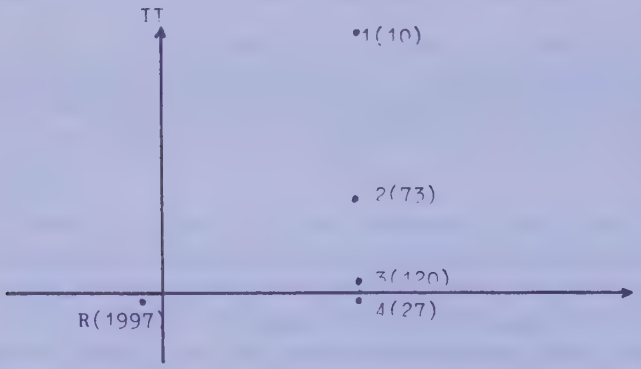


Abb. 4.10: Trennung zwischen Studierenden der Math/Inf mit den Vordiplomzensuren 1,2,3 und 4 und den Studierenden der anderen Fächer durch die Abiturnoten

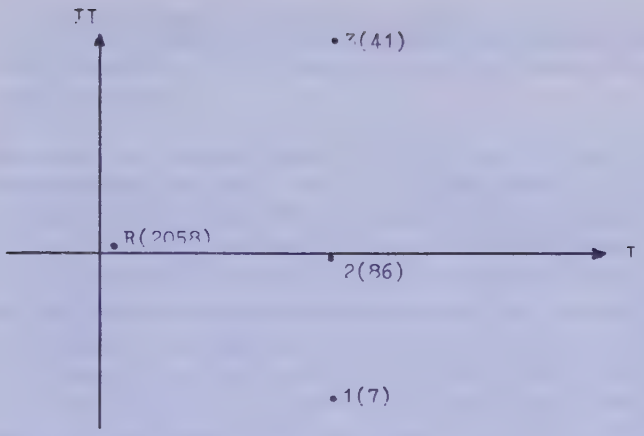


Abb. 4.11: Trennung zwischen Studierenden des Faches Deutsch mit den Zensuren im Zwischenexamen 1,2 und 3 und den Studierenden der anderen Fächer durch die Abiturnoten

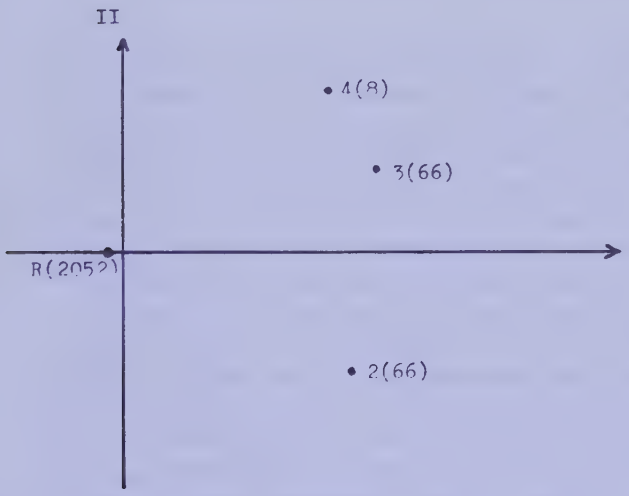


Abb. 4.12: Trennung zwischen Studierenden des Faches Französisch mit den Zensuren im Zwischenexamen 2,3 und 4 und den Studierenden der anderen Fächer (R) durch die Abiturnoten

Tabelle 4.4: Die von den rotierten Diskriminanzfunktionen aus den Prädiktoren extrahierten Varianzanteile

	Die Studienfächer unterscheidende Diskriminanzfunktion	Die Erfolgsgruppen unterscheidende Diskriminanzfunktion	Beide Funktionen zusammen
BWL/VWL	13,5 %	19,4 %	32,9 %
Medizin	15,1 %	24,7 %	39,8 %
Math/Inf	5,5 %	12,5 %	18 %
Deutsch	10,8 %	27,3 %	38,1 %
Französisch	8,9 %	29,4 %	38,6 %

Bei der zweidimensionalen Darstellung der zu trennenden Gruppen (vgl. Abb.4.8 - 4.12) kann man zwei Typen von Studienfächern unterscheiden:

1. Bei den Studienfächern Math/Inf, Deutsch und Französisch werden dem Fach angehörende und nicht dem Fach angehörende Personen durch eine bipolare Komponente, die sowohl positiv als auch negativ auf bestimmten Abiturnoten lädt (vgl. auch 4.2) getrennt. Dagegen werden die Gruppen der verschiedenen Examenszensuren wie bei allen Fächern durch eine unipolare, generelle Komponente der Abiturnoten getrennt (vgl. auch 3.4.1). Bei dem unterschiedlichen Charakter beider Komponenten ist es einleuchtend, daß sie bei gleichzeitiger Betrachtung in der vorliegenden Analyse unvermischt blieben. Dadurch war es ohne bzw. mit geringfügiger Rotation möglich, die Zensurengruppen jeweils nur durch die eine, und die Fachunterscheidung jeweils nur durch die andere Komponente darzustellen (Abb. 4.10 - 4.12).

2. Bei BWL/VWL und Medizin wird die Fachunterscheidung (vgl. auch 4.2) ebenso wie die Unterscheidung der Erfolgsgruppen (vgl. auch 3.4.1) durch unipolare, generelle Komponenten der Abiturnoten getroffen, so daß eine Vermischung beider Dimensionen in der vorliegenden gemeinsamen Analyse zu erwarten war. Bei Medizin ließ sich durch Rotation um ca. 45^{o} eine klare Trennung der beiden Komponenten erreichen. Dabei liegt die Restgruppe der nicht zum Fach gehörenden Personen in bezug auf die die Zensuren unterscheidende Dimension nicht in der Mitte, sondern nahe der schlechtesten Zensurengruppe (Abb. 4.9). Bei BWL/VWL war eine Trennung der beiden Dimensionen auf fach- und zensurenunterscheidend auch nach Rotation um ca. 45^{o} nur in etwa zu erreichen (Abb. 4.8). Hier liegt die Restgruppe ebenfalls nicht in der Mitte der die Zensuren unterscheidenden Dimension, sondern nahe der besten Zensurengruppe.

Insgesamt lassen sich also bei allen Studienfächern - etwas eingeschränkt nur bei BWL/VWL - nach visueller Rotation zwei Dimensionen unterscheiden, die jeweils nahezu rein einmal die Fachzugehörigkeit und zum anderen die Zensur im Zwischenexamen des untersuchten Fachs repräsentieren.

Der Anteil an Varianz, den diese Komponenten innerhalb der Abiturnoten umfassen, kann der Tab. 4.4 entnommen werden. Dabei muß allerdings die Menge der extrahierten Varianz keineswegs mit der Stärke der Vorhersage der Kriterien zusammengehen. Z.B. extrahiert die studienfachunterscheidende Funktion bei Math/Inf nur 5,5 % der Varianz der Abiturnoten, obwohl die Vorhersage gerade dieses Studienfachs (vgl. Tab. 4.3) besser als bei allen anderen ist. Offenbar war hier die wenige - besonders aus der Mathematiknote - herausgezogene Varianz besonders wirksam für die Vorhersage. Bei Medizin dagegen wird z.B. mit wesentlich mehr Varianz der Abiturnoten (15 % laut Tab. 4.4) eine deutlich geringere Vorhersage des Studienfachs nach Tab. 4.3 erreicht.

Kann man nun behaupten, daß die Vorhersage des Studienerfolgs durch die Abiturnoten besser wäre, wenn vorher bei der Wahl des Studienfachs nicht bereits eine Auslese nach Abiturnoten stattfände? Im direkten empirischen Angriff läßt sich diese Frage nicht entscheiden, weil sich die dazu notwendige Situation - Zuteilung der Studienfächer an die Abiturienten nach Zufall -aus vielerlei Gründen nicht herstellen läßt. Wohl aber kann man aus dem bisher Festgestellten einige Hinweise für die Beantwortung dieser Frage entnehmen. Es ist plausibel, daß jener erste Anteil der Abiturnoten, der das Studienfach vorhersagt, und somit für die Vorhersage des Erfolgs "verbraucht" ist, im (irrealen) Fall einer zufallsmäßigen Zuteilung der Studienfächer ebensogut den Studienerfolg vorhersagen könnte wie die zweite Komponente. Nicht nur stünde dann überhaupt mehr Varianz der Abiturnoten - vielleicht etwa die in der letzten Spalte der Tab. 4.4 genannten Anteile - zur Vorhersage des Erfolgs zur Verfügung, sondern es ist darüber hinaus plausibel,

daß der Anteil der Abiturnoten, der hier das Studienfach vorhersagt, eigentlich für die Vorhersage gerade des Studienerfolgs besonders valide wäre. Zum einen handelt es sich nämlich bei den Anteilen der Abiturnoten, die das Studienfach vorhersagen können, besonders um jene Schulfächer, die wir als die dem Studienfach homologen Schulfächer nach allem Wissen als die validesten Prädiktoren zur Vorhersage des Studienerfolgs ansehen müßten. Zum anderen dürften sich auch die Abiturienten bei ihrer Studienfachwahl stark an solchen Merkmalen orientieren, von denen sie - sicherlich zu Recht - einen Zusammenhang mit dem späteren Studienerfolg annehmen. Dabei haben die Abiturienten zusätzlich die Möglichkeit, diese Merkmale nach Art und Weise ihres Zustandekommens zu korrigieren und zu gewichten. Z.B. können die Abiturienten bestimmte Abiturzensuren, die wesentlich vom persönlichen Verhältnis zum betreffenden Lehrer beeinflußt sind, ähnlich dem in 3.4.1 gefundenen Suppressoreffekt dieses Merkmals korrigieren. U.a. in solchen Korrekturprozessen könnte auch die Ursache dafür liegen, daß sich in Kap. 3.4.1 die relativ niedrige Validität der den Studienfächern korrespondierenden Abiturnoten im wesentlichen nicht durch eine verringerte Streuung dieser Abiturnoten erklären ließ. D.h. die "Rest"-Varianz der homologen Abiturfächer, mit denen in 3.4.1 der Studienerfolg vorherzusagen versucht wurde, besteht im wesentlichen aus relativ irrelevanter Varianz, die nach Anwendung von ähnlichen, wie den oben angedeuteten Korrekturmaßnahmen übrig blieb.

Es ist also gut möglich, daß die Komponente der Abiturzensuren die das Studienfach vorhersagt, trotz eines relativ geringen Varianzanteils - theoretisch - eine besonders hohe Validität zur Vorhersage des Studienerfolgs besitzt. Insofern kann man annehmen, daß die "wahre" Validität der Abiturnoten insgesamt wesentlich größer als die in der Wirklichkeit gefundene ist. Darüber hinaus sind natürlich noch weitere, häufiger genannte Einflüsse wirksam, die die empirisch bestimmte Validität der Abiturnoten reduzieren. Als Beispiele seien hier die meist

nicht berücksichtigte Auslese der Studenten durch vorzeitigen Studienabbruch, sowie die unzureichende Standardisierung der Abiturnoten (z.B. INGENKAMP, 1976) genannt.

4.4 Vorschlag eines Modells, das gleichzeitig Studienerfolg und Studienabbruch als Kriterien einbezieht

Analog zu dem in 4.3 behandelten Versuchsaufbau läßt sich ein Plan entwerfen, der gleichzeitig die Examenszensuren und den möglichen Abbruch des betr. Fachs (und vielleicht weitere, ähnliche Merkmale) als Gruppenmerkmale in eine Diskriminanzanalyse, vielleicht auch eine faktorielle Diskriminanzanalyse (vgl. COOLEY und LOHNES, 1971, Kap. 12) einbezieht. Ein solcher möglicher Plan wird in Abb. 4.13 schematisch veranschaulicht.

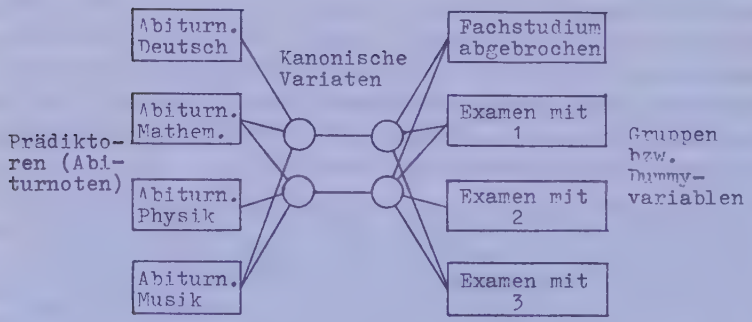


Abb. 4.13: Die Vorhersage des Studienerfolgs (Studienabschluß und Examenszensur) durch die Abiturnoten mit Hilfe einer Diskriminanzanalyse

Dabei wäre zu prüfen, wieviele Diskriminanzfunktionen anzunehmen sind. Möglicherweise wäre nur eine Funktion vorhanden, d.h. die Abbrecher könnten auf derselben Funktion abgebildet werden wie die Gruppen mit unterschiedlicher Examenszensur. Durch die Hinzufügung der Abbrecher in diesem Plan wäre dann die eine Diskriminanzfunktion vermutlich weiter aufgespreizt als vorher. Dabei würde sich ein höherer Zusammenhang zwischen Prädiktoren und einem quantitativen Kriterium in einer multiplen Regressionsanalyse ergeben, wenn man die Abbrecher mit der entsprechenden, aus der Diskriminanzanalyse ermittelten "Zensur" der normalen Zensurenskala anfügen würde, um ein neues, zusammengesetztes Kriterium zu definieren. In diesem Fall wären die z.B. in 3.4 gefundene Auswahl der Prädiktoren und ihre Gewichtung auch für die Vorhersage des Studienabbruchs gültig, würden also für ein breiteres Kriterienspektrum gelten als zunächst erwartet.

Die andere - wohl wahrscheinlichere - Möglichkeit ist, daß wenigstens zwei Dimensionen zur Vorhersage der Kriterien nötig sind. Dann würde die eine Dimension zwischen den verschiedenen Examenszensuren und die andere zwischen den Abbrechern und den erfolgreich Studierenden differenzieren. In diesem Fall würden also Examenszensuren und Studienabbruch von unterschiedlichen Komponenten der Abiturzensuren vorhergesagt,und die bisher gefundenen Vorhersagegleichungen für die Examenszensuren wären nicht auf das Kriterium Studienabbruch übertragbar.

4.5 Zusammenfassung

Das Kriterium "Zensur im Zwischenexamen" wurde auf Gleichabständigkeit der Skalenstufen und Eindimensionalität indirekt in Form der Hypothese geprüft, daß sich das Kriterium unter diesen Voraussetzungen optimal durch die Prädiktoren abbilden läßt. Konkret wurde gefragt, ob von einer Diskriminanzanalyse mit dem Kriterium als Gruppenmerkmal die zusätzlichen Freiheitsgrade durch eine bessere Anpassung des Modells an die Daten genutzt wurde. In den Ergebnissen ließ sich das Kriterium eindeutig eindimensional darstellen; die Gleichabständigkeit der Stufen des Kriteriums erschien in den Ergebnissen hinreichend gut angenähert. Somit ist also die Regressionsanalyse gerechtfertigt, und sie stellt als das sparsamere und effizientere Modell die beste Wahl des Verfahrens dar.

Die Unterscheidung der Studienfächer aufgrund der Abiturnoten ließ sich in zwei Dimensionen darstellen. Bei einer (Re-) Klassifikation der Vpn wurde eine signifikante Zuordnungsquote erreicht. Die erste rotierte Dimension, gekennzeichnet durch gute Mathematikzensuren gegenüber schlechteren Zensuren in den sprachlichen Fächern, trennte vor allem zwischen den Studienfächern Math/Inf einerseits und den sprachlichen Fächern, sowie Medizin andererseits. Die zweite rotierte Dimension, gekennzeichnet durch generell schlechtere Noten vor allem in Physik, Englisch und Biologie trennte BWL/VWL von den restlichen Studienfächern. Den günstigsten Zensurendurchschnitt auf dieser Dimension wiesen Medizin und Math/Inf auf. Wurden jeweils die einzelnen Studienfächer der Gesamtheit der restlichen Fächer gegenübergestellt, so bestätigte sich dieses Bild bei feinerer Beschreibung der Gegebenheiten des jeweils betrachteten Fachs.

Schließlich wurde die gleichzeitige Vorhersage der Kriterien "Zensur im Zwischenexamen" und "Wahl des Studienfachs" mit Hilfe von Diskriminanzanalysen untersucht. Dabei konnten die

genannten Kriterien jeweils durch eine von zwei orthogonalen Komponenten der Abiturnoten vorhergesagt werden. Jede dieser beiden Komponenten zog einen beträchtlichen Teil der gesamten Varianz der Abiturnoten auf sich. Es wurde argumentiert, daß der Anteil der Abiturnoten, der das Studienfach vorhersagt und somit für die Vorhersage des Studienerfolgs "verbraucht" ist, eigentlich, d.h. im gedachten Fall einer zufallsmäßigen Zuteilung der Studienfächer an die Abiturienten, besonders gut zur Vorhersage gerade des Studienerfolgs geeignet ist. Insofern muß man die (theoretisch) "wahre" Validität der Abiturnoten deutlich höher und mit ihrem Schwerpunkt bei anderen Schulfächern liegend annehmen als die empirisch bestimmte Validität.

Ein ähnliches, wie das den vorstehenden Ausführungen zugrunde liegende Modell wurde zur Analyse von Daten vorgeschlagen, die als Kriterien gleichzeitig "Examenszensuren" und "Studienabbruch" enthalten. Dazu werden die genannten Kriterien gleichzeitig als Gruppenmerkmale in Diskriminanzanalysen untersucht. Zwei Typen von möglichen Ergebnissen, nämlich die Darstellbarkeit der Abbrecher auf der gleichen Dimension wie die Examenszensuren oder auf einer anderen, weiteren Dimension, wurden im Hinblick auf die Bewertung der bei der Vorhersage der Examenszensuren allein erzielten Ergebnisse besprochen.

5. Erweiterungen des linearen Modells

Dieses Kapitel wird mit Abweichungen von den Voraussetzungen des linearen Modells befaßt sein, und es soll versucht werden, die Struktur der Daten mit Hilfe von Modifikationen des linearen Modells aufzudecken. Dabei wird - ebenso wie in allen folgenden Kapiteln - als Kriterium wieder die "Zensur im Zwischenexamen" fungieren.

Es sollen vor allem folgende Ansätze erprobt werden:

- Analyse von Wechselwirkungen zwischen den Prädiktoren nach Art der Wechselwirkungen bei einer Varianzanalyse. Dies entspricht einer Moderatoranalyse nach SAUNDERS (1956).
- Analyse der Residuen der Regressionsrechnungen von 3.3 zur Entdeckung von nichtlinearen Effekten und zur Überprüfung von Modellannahmen (vgl. z.B. ANSCOMBE und TUKEY, 1963).
- Analyse der absoluten Differenzen zwischen standardisierten Schätz- und Kriteriumswerten nach dem Ansatz der Moderatoranalyse von GHISELLI (1956).

Die Moderatoranalyse, die z.B. auch von MEREDITH et al. (1974) gemeinsam mit nichtlinearer Regression und "stückweiser" (piecewise) Regression, d.h. dem Arbeiten mit Residuen, behandelt wird, ist eher als logisches, theoretisches Konzept statt als technisches Verfahren aufzufassen. Ansätzen der Moderatoranalyse wird in den letzten ca. 10 Jahren große Aufmerksamkeit zugewendet. Gelegentlich wird jedoch beklagt, daß unter dieser Bezeichnung verschiedenste Techniken zusammengefaßt werden, daß es noch keine allgemein akzeptierten Techniken gibt und daß die Diskussion über Einsatzfähigkeit und Effektivität der einzelnen Techniken keineswegs abgeschlossen ist. Deshalb soll zunächst im folgenden Abschnitt versucht werden, eine - wenn auch sehr kurze und sicherlich unvollständige - Übersicht über wichtigste Techniken zu geben und die wesentlichen Linien für die vorliegende Untersuchung herauszuheben.

5.1 Moderatoranalyse

Die Wurzeln und zugleich die wichtigsten Richtungen der Moderatoranalyse finden sich bei SAUNDERS (1955) und bei GHISELLI (1956). Die Bezeichnung Moderator wurde von SAUNDERS eingeführt, von GHISELLI später übernommen und schließlich allgemein akzeptiert. Daneben finden sich aber immer noch Bezeich-

nungen (nach ZEDECK, 1971 a) wie "population control variable", "subgrouping variable", "predictability variable" (nach GHISELLI, 1956) "homologizer variable" u.a. Gemeinsam ist all diesen Bezeichnungen und Konzepten, daß eine Vorhersage durch einfache oder multiple Korrelation mit Hilfe von Moderatorvariablen verbessert wird, die nicht der Einbeziehung des Moderators als einfachem Prädiktor in eine multiple Korrelation entspricht.[1] Eine starke Motivation zur Anwendung von Moderatortechniken entstammte dem Befund von GHISELLIS (1966) Übersicht, daß bei Prädiktionsstudien allgemein nicht mehr als maximal 25 % der Varianz des Trainingserfolgs oder 12 % des Berufserfolgs erklärt wurden. RUNDQUIST (1969), der verschiedene, zur Überwindung dieser "prediction ceiling" geeignete Verfahren bespricht, kommt zu dem Schluß, daß nur der Ansatz der Moderatoranalyse nach wie vor erfolgversprechend erscheint. Dagegen hätten sich andere Verfahren wie Faktoren- und Itemanalyse - beides lineare Modelle - als unfähig zur Überwindung dieser Barriere erwiesen. Er plädiert dafür, Vorhersage innerhalb offener Systeme zu betreiben und die dynamische Fähigkeit der beteiligten Variablen zur Wechselwirkung miteinander stärker zu beachten.

Nachfolgend sollen die drei wichtigsten Richtungen der Moderatoranalyse kurz beschrieben werden, wobei in etwa die Einteilung von ZEDECK (1971 a) verwendet wird:

1. Der älteste Ansatz, der nicht unbedingt mit einem Moderatorkonzept verbunden sein muß, ist die Analyse von Subgruppen, die sich durch ihre Ausprägung auf einer Moderatorvariable unterscheiden. Es wird geprüft, ob sich diese Subgruppen in unterschiedlicher Weise vorhersagen lassen. Ein bekanntes Beispiel ist die Vorhersage des Studienerfolgs getrennt für

---

1) Häufig beobachtet man nämlich, daß beim Aufbau einer multiplen Regressionsgleichung von einem bestimmten Punkt an keine weitere Steigerung mehr möglich ist, obwohl noch viele interessante Prädiktoren nicht in die Gleichung aufgenommen sind.

Studenten und Studentinnen, wobei sich in den USA die Studentinnen konsistent als besser vorhersagbar erwiesen (z.B. GROSS et al., 1974 b). Dieser Ansatz, auch als "heterogeneous regression" bezeichnet, wird z.B. in neuerer Zeit von VELICER (1972a) beschrieben und propagiert. Er scheint bis jetzt auch das weitaus häufigste Verfahren bei anwendungsorientierten Untersuchungen zu sein, wie z.B. bei einer Durchsicht der letzten Jahrgänge des Journal of Applied Psychology festgestellt wurde. Einschneidende Nachteile dieses sonst sehr einleuchtenden und leicht zu durchschauenden Verfahrens liegen in der Notwendigkeit, von vornherein einigermaßen aussichtsreiche Moderatorvariablen und sehr umfangreiche Stichproben, die die entsprechende Unterteilung zulassen, zur Verfügung zu haben.

2. Die Methode von SAUNDERS (1956) - in Deutschland von BARTUSSEK (1970) vorgeschlagen - erlaubt die Aufstellung einer für die ganze Stichprobe gültigen Vorhersagegleichung. Hauptmerkmal der auch als "moderated regression" bezeichneten Methode ist, daß die Gewichte der Prädiktoren in der Vorhersagegleichung nicht konstant, sondern eine lineare Funktion von Moderatorvariablen sind. Dabei können die Moderatoren nicht nur diskrete Variablen wie bei Methode 1, sondern auch kontinuierliche sein.

   Dieses Konzept führt zu einer Regressionsgleichung, die eine Linearkombination nicht nur der Ausgangsvariablen, sondern auch der Produkte von Prädiktoren und Moderatoren enthält:

$$y = \bar{y} + a_i x_i + b_j z_j + c_{ij} x_i z_j$$

   Dabei sind a, b und c Gewichte, x Prädiktoren, z Moderatoren und y das Kriterium. Das Modell bleibt linear und mit den gewohnten Methoden analysierbar, wenn man die Produktvariablen $x_i z_j$ über eine Variablentransformation als neue Variable $w_{ij} = x_i z_j$ definiert:

$$y = \bar{y} + a_i x_i + b_j z_j + c_{ij} w_{ij}$$

Gemessen an den Ausgangsvariablen wird allerdings der Rahmen des linearen Modells mit der Nichtadditivität der Effekte gesprengt. Mit dieser Methode gelang es, unter Verwendung des Merkmals Zwanghaftigkeit als Moderator die Vorhersage der akademischen Leistung zu verbessern, obwohl Zwanghaftigkeit als einfacher Prädiktor keinerlei Vorhersageeffekt hatte (SAUNDERS, 1953 und MELVILLE, 1954, zit. nach ROMINE et al. 1970). Die Einführung von Produktvariablen in die Regressionsgleichung entspricht der Einführung von Wechselwirkungen in die Varianzanalyse und ist in den Standardtexten beschrieben (z.B. KERLINGER et al., 1973, DRAPER et al., 1966). Diese Methode soll in 5.2 erprobt werden.

3. Diese Richtung der Moderatoranalyse - vertreten vor allem durch GHISELLI (1956 und viele andere Arbeiten) - hat wohl die meisten Anwendungen und Modifikationen der Methode angeregt und ist wohl auch am häufigsten mit praktischem Erfolg verbunden gewesen (CONGER, 1969). Kernstück der wichtigsten Technik (GHISELLI, 1960) ist die Analyse der absoluten Differenzen zwischen standardisierten Schätz- und Kriteriumswerten als Maß für das bessere oder schlechtere Zutreffen der Regressionsgleichung bei verschiedenen Gruppen von Individuen. Es werden Merkmale gesucht, die mit diesen Differenzen korreliert sind, und die gesamte Stichprobe wird aufgrund einer solchen Moderatorvariable in zwei Teile zerlegt. Im einen der so gebildeten Teile der Stichprobe arbeitet dann die gefundene Gleichung zufriedenstellend; im anderen Teil, in dem die Gleichung weniger gut arbeitet, muß nach anderen Möglichkeiten der Vorhersage gesucht werden.

   Dieser Ansatz zeigt deutlich eine Veränderung der klassischen Theorie der Vorhersage: Die klassische Theorie setzt sowohl bei der Messung (Reliabilität) als auch bei der Vorhersage (Validität) gleiche Erwartungswerte der Fehler bei allen Individuen voraus. Erfolgreiche Anwendung von GHISELLIs Methode

zeigen jedoch, daß es Variablen (Moderatoren) gibt, die die Abweichungen der wirklichen von den geschätzten Kriterienwerten vorhersagen können, d.h. daß die Vorhersagefehler systematisch zwischen den Individuen variieren (GHISELLI, 1963). Dieses Moderatorkonzept kann sogar auch dann gültig sein, wenn sich empirisch Homoskedastizität der Fehler bei der Vorhersage eines Kriteriums nachweisen läßt. Es ist nämlich möglich, daß die beobachtete Homoskedastizität durch die Überlagerung von Untergruppen mit verschiedenen Arten der Heteroskedastizität zustandekommt (GHISELLI, 1967).

Insgesamt gesehen, scheint die Bereitschaft der Forscher, das klassische Vorhersagemodell durch das Moderatorkonzept zu erweitern und zu einem offeneren System überzugehen, groß zu sein. Erschwerend und hindernd für eine praktische Anwendung der Konzepte auf breiter Basis nicht nur durch methodisch orientierte Untersucher bleibt jedoch die Tatsache, daß eine ganze Reihe verschiedener Techniken ohne klare und allgemein akzeptierte Regeln für ihre Indikation nebeneinander besteht. Als Beispiele für weitere Techniken sind zusätzlich zu den vorstehend beschriebenen Methoden zu nennen:

GHISELLIs System für zwei Prädiktoren (GHISELLI, 1960 und 1963), mathematisch formalisiert mit Mißverständnissen der Methode GHISELLIs - vgl. ABRAHAMS et al., 1972 a - von CONGER (1969),

Die Quadrantenanalyse von HOBERT und DUNNETTE (1967),

Das "multipredictable group validation model" von ZEDECK (1971 b),

Die Methode der multiplen Moderatoren (ROCK et al., 1967, KLEIN et al., 1968), bei der das Verfahren der hierarchischen Gruppierung von WARD (1963) zur Bildung von Untergruppen eingesetzt wird.

Es muß beachtet werden, daß in einer Phase der regen Entwicklung neuer Techniken - die (scheinbar) klare und einfache Situation z.B. bei der Analyse von Differenzwerten fordert offensichtlich zur Entwicklung und Modifizierung von Techniken heraus - leicht

Mißverständnisse und Fehler auftauchen. Kritische Bemerkungen zu verschiedenen vorgeschlagenen Techniken bringen die unten aufgezählten Arbeiten, die jeweils zur Illustration des Inhalts und Anlasses mit ihrem Titel aufgeführt sind. U.a. werden durch die darin enthaltenen Kritiken GHISELLIs System für zwei Prädiktoren und die Quadrantenanalyse wenigstens vorerst in ihrer Anwendbarkeit wesentlich eingeschränkt.

McNEMAR (1969): Moderation of a Moderator Technique

ABRAHAMS (1969): Off Quadrant Comments

VELICER (1972): Comment on the General Inapplicability of Ghiselli's Moderator System for two Prediktors

ABRAHAMS and ALF (1972 a): Pratfalls in Moderator Research

DUNNETTE (1972): Comment on Abrahams and Alf's "Pratfalls in Moderator Research"

ABRAHAMS und ALF (1972 b): Reply to Dunnete's "Comment on Abrahams and Alf's "Pratfalls in Moderator Research""

Ein wichtiges grundsätzliches Problem stellt schließlich die Prüfung dar, ob bei der Anwendung einer Moderatortechnik auf Auslesesituationen wirklich ein praktischer Vorteil erzielt wird. So reicht es keineswegs aus, eine Erhöhung der multiplen Korrelation durch Kreuzvalidierung nachzuweisen. Es ist immerhin möglich, daß die Korrelationen durch die Vergrößerung der Streuungen von beteiligten Merkmalen anstiegen, während das Ausmaß der Fehlervarianz als wichtigerer Indikator für die Güte der Vorhersage konstant blieb. PINDER (1973) weist darauf hin und zeigt, daß ein Gewinn an multipler Korrelation leicht durch den mit Anwendung einer Moderatortechnik eingetretenen Verlust an statistischer Effizienz ausgewogen werden kann. Von GROSS et al. (1974)

werden verschiedene statistische Verfahren zur Bewertung der Nützlichkeit der mit Hilfe einer Moderatortechnik erzielten Ergebnisse vorgeschlagen. Die Auswirkungen von Moderatortechniken auf den Ausgleich der Chancen von unterprivilegierten Personengruppen werden schließlich von JONES (1973) diskutiert.

Die Bedeutung, die dem Moderatorkonzept als einem das rein lineare Modell verlassenden Konzept beigemessen wird, ist richtungsweisend auch für die vorliegende Untersuchung. Es werden die beiden wichtigsten, in den vorstehenden Ausführungen unter Punkt 2 und 3 aufgeführten Techniken, die auch nach den verschiedenen Krtiken an Moderatortechniken unbestritten bleiben und sich theoretisch überzeugend darstellen, in 5.2 und 5.4 durchgeführt. Eine Auswertung des Moderatoransatzes auf praktische Anwendung hin wird lediglich für den Fall in Aussicht genommen, daß Moderatoreffekte von größerer Bedeutung gesichert werden können.

Außerdem werden in 5.3 Residuen analysiert; der Ansatz entspricht teilweise dem als "piecewise regression" bezeichneten (z.B. MEREDITH et al., 1974), bei dem eine zweite bzw. weitere Gruppe von Prädiktoren jeweils mit den Residuen, die aus einer ersten bzw. vorhergehenden Gruppe von Prädiktoren gewonnen wurden, in Beziehung gesetzt wird.

Die in 5.3 und 5.4 durchgeführten Verfahren beziehen sich auf ähnliche Merkmale: In 5.4 handelt es sich um absolute Differenzen zwischen standardisierten Schätz- und Kriteriumswerten, in 5.3 um algebraische Differenzen zwischen unstandardisierten Schätz- und Kriteriumswerten.

5.2 Moderatoranalyse nach SAUNDERS

Wegen der großen Zahl möglicher neuer, durch Transformation entstehender, Prädiktoren wurden nur die drei größten Studienfächer BWL/VWL, Medizin und Math/Inf analysiert. Gestützt auf bekanntgewordene Ergebnisse und persönliche Vermutungen wurde folgende Liste von "moderatorverdächtigen" Variablen aufgestellt, die in diesem Abschnitt verwendet und auch in weiteren Verfahren berücksichtigt werden soll:

Geschlecht; wie bereits erwähnt, wurde dieses Merkmal in den USA häufig als Moderator identifiziert,

Schicht; die Bedeutung dieses Merkmals illustriert z.B. SELVINs Behauptung (1969), daß die meisten Studien der Struktur von Einstellungen und ähnlicher Konstrukte lediglich Scheinbeziehungen aufgrund von unkontrollierter Streuung der sozialen Schicht und anderer fundamentaler sozialer Merkmale aufgedeckt haben,

Depression,

Verhältnis zu den Lehrern,

Fachinteresse,

Prüfungsangst,

die dem Studienfach korrespondierende Abiturnote,

die in Kap. 3.3 ausgewählten "harten" Prädiktoren.

Es wurden zunächst alle der sich aus den aufgeführten Variablen ergebenden Produkte bis zur höchsten Ordnung berechnet; ausgenommen waren lediglich die Produkte von Abiturnoten untereinander. Wegen der großen Anzahl der Effekte wurde sodann nach bestimmten Vorüberlegungen ein Teil der möglichen Produkte - vor allem höherer Ordnung - ausgeschieden, so daß sich endgültig für BWL/VWL 20, für Medizin 20 und für Math/Inf 28 Produktvariablen ergaben. Die Prädiktoren wurden schrittweise nach einer festen Reihenfolge in die Regressionsanalyse

eingeführt (vgl. 3.3); die Reihenfolge ging von den Haupteffekten, über die Wechselwirkungen (= Produktvariablen) niedriger zu den Wechselwirkungen höherer Ordnung.

Die Ergebnisse waren eindeutig; auch nicht in Tendenzen zeigten sich Effekte der Produktvariablen. Mit der Methode von SAUNDERS ließen sich also bei den hier verwendeten Variablen und Variablenkombinationen keine Moderatoreffekte nachweisen. Dieses Ergebnis muß jedoch keineswegs die Abwesenheit jeglicher Moderatoreffekte bedeuten. In einem Vergleich verschiedener Moderatortechniken versagte z.B. die Methode von SAUNDERS dabei, bereits mit anderen Techniken entdeckte Moderatoren zu identifizieren (ZEDECK, 1971 c). Der gleiche Autor (ZEDECK, 1971 a) berichtet von zwei weiteren Studien mit ähnlichen Ergebnissen, und er stellt fest, daß kaum jemals mit der Methode der moderierten Regression erzielte positive Ergebnisse bekannt wurden. Möglicherweise stellt diese theoretisch elegante und effiziente Methode formal zu strenge Anforderungen an die Daten: Die Moderatorvariable muß intervallskaliert und kontinuierlich sein, und sie muß in sehr regelhafter Art mit den Prädiktoren interagieren. ZEDECK vermutet (1971 c), daß man möglicherweise durch Erhebung der Prädiktoren und Moderatoren in höhere Potenzen mehr Erfolg haben könnte. M.E. ist dieses Verfahren jedoch wegen der außerordentlich großen Zahl möglicher Kombinationen ohne klare Hypothesen a priori kaum anwendbar.

5.3 Analyse der Residuen

Auch hier wurden außer der Gesamtgruppe der Studenten mit fachweise standardisiertem Kriterium nur die drei größten Studienfächer berücksichtigt. Es wurden jeweils Residuen unter Auspartialisierung der drei in Kap. 3.3 unterschiedenen Prädiktorengruppen (Abiturnoten, "harte" und "alle" Prädiktoren) berechnet. Das Ziel der anschließenden Prüfungen war es, Aufschlüsse über nichtlineare Effekte von Prädiktoren in diesem Stadium - und nicht wie in 3.2.2 vor Beginn der Analyse überhaupt -, sowie über die Angemessenheit des sta-

tistischen Modells zu erhalten. Außerdem sollten hier auch die Merkmale "Studiendauer" und "Prüfungsangst" in die Betrachtung einbezogen werden. Zur Analyse der Residuen wurden vor allem beschreibend, teils auch inferentiell statistische Methoden eingesetzt.

Zunächst wurde die Gesamtgruppe, die Angehörige verschiedener Studienfächer enthielt, auf die in diesem Fall von vornherein angezweifelte Voraussetzung der unabhängigen Fehler untersucht. Ein nonparametrischer Runs Test auf Abfolge positiver und negativer Residuen (vgl. z.B. JOHNSTON, 1963), der sich bereits innerhalb des verwendeten Rechenprogramms durchführen ließ, hatte lediglich bei der Gesamtgruppe ohne Standardisierung des Kriteriums mit einer Signifikanz angesprochen und damit die Notwendigkeit einer fachweisen Standardisierung des Kriteriums unterstrichen. Wegen der relativ geringen Effizienz des Runs-Tests, und weil auch die Anordnung der Vpn in der Datei für die Entdeckung der zu prüfenden Fakten nicht optimal einheitlich nach Fächern angelegt war, wurde zusätzlich der direkte Weg zur Untersuchung der Abhängigkeit der Residuen von der Fachzugehörigkeit in Form von bivariaten Häufigkeitstabellen beschritten. Als Mittelwerte der Residuen nach Auspartialisierung der Abiturnoten - stellvertretend für die gleichsinnigen Befunde bei den anderen Prädiktorgruppen - ergaben sich für die Angehörigen der Studienfächer folgende Zahlen:

BWL/VWL	- 0,59
Medizin	0,19
Math/Inf	0,93
Deutsch	0,36
Französisch	- 0,22
Chemie	0,18
Wirtschaftspädagogik	- 0,35
Physik	0,46

Das bei der Vorhersage der Residuen durch die Fachzugehörigkeit zu erzielende Eta-Quadrat betrug 0,122, d.h. 12 % der Varianz der Residuen ist durch die Fachzugehörigkeit festgelegt ($p < 0{,}001$), obwohl das Kriterium fachweise standardisiert worden war. Entsprechende Ergebnisse traten auch bei fachweiser Standardisierung der

Prädiktoren auf. Es werden die Studenten der Math/Inf, Physik und Deutsch über-, der BWL/VWL und Wirtschaftspädagogik systematisch unterschätzt. Die gefundene Regressionsgleichung wird also den Angehörigen der Fächer nicht in gleicher Weise gerecht; die Fehler sind fachweise miteinander korreliert.

Des weiteren wurden an den Stichproben der drei Studienfächer BWL/VWL, Medizin und Math/Inf folgende Berechnungen durchgeführt:

- Bivariate Häufigkeitstabellen der Residuen mit allen anderen Variablen und den Schätzwerten
- Eta-Quadrat- und Korrelationskoeffizienten zu den vorstehend genannten Kreuztabellen
- Mittelwerte und Streuungen der Residuen innerhalb der Stufen von ausgewählten, interessant erscheinenden Variablen.

Um die Stärke der linearen oder nichtlinearen Vorhersage einzelner, nicht als Prädiktoren ausgewählter Variablen beurteilen zu können, wurden sie mit solchen Prädiktoren verglichen, die im jeweiligen Stadium zwar noch nicht, wohl aber innerhalb einer erweiterten Gruppe von Prädiktoren (z.B. beim Übergang von den "harten" auf "alle" Prädiktoren) ausgewählt wurden.[1)] Dadurch konnte ein ungefährer Anhaltswert dafür gewonnen werden, wann ein Prädiktor, der einen bestimmten Zusammenhang mit den Residuen hatte, tatsächlich einen Beitrag zur multiplen Vorhersage bringen würde. Der vielleicht signifikante Zusammenhang eines Merkmals mit den Residuen allein kann nämlich keineswegs - wie es vielleicht vordergründig erscheinen möchte - als Gewährleistung für einen Beitrag dieses Merkmals zur multiplen Vorhersage genommen werden. Denn die Residuen können nicht als gewöhnliche Meßwerte aufgefaßt werden, sondern sie sind mit allen statistischen Unsicherheiten der vorangegangenen Rechnungen behaftet. Der Beitrag eines Merkmals zur multiplen Vorhersage kann also nur innerhalb des Gesamtsystems beurteilt werden.

---

1) Hilfreich war dabei zusätzlich die Betrachtung der nach Abbruch der Regressionsanalysen nächststarken Prädiktoren, die also "beinahe" in die Regressionsgleichung aufgenommen worden wären.

In bezug auf die Qualität der Zusammenhänge wurden - ähnlich wie in 3.2 - durch Vergleich von Eta-Quadrat- und Korrelationskoeffizienten Schlüsse auf die Stärke der linearen und nichtlinearen Komponenten gezogen.

Anschließend folgt eine Aufzählung der wichtigsten Ergebnisse; nebensächlichere Befunde wie z.B. Bestätigungen des Modells, der Prädiktorenauswahlen und der Richtigkeit der Rechnungen werden nicht aufgeführt:

1. Es gab keine Hinweise auf kurvilineare Effekte von solcher Stärke, daß sich eine Berücksichtigung innerhalb der Regressionsanalysen lohnen würde.

2. Es wurde der Zusammenhang des Merkmals "Studiendauer" mit den unter Auspartialisierung "aller Prädiktoren" gewonnenen Residuen untersucht. Bei den Fächern BWL/VWL und Math/Inf wurde ein linearer Trend - kürzeres Studium geht mit besseren Zensuren zusammen - beobachtet. Der häufig berichtete Zusammenhang zwischen Studiendauer und Zensur verschwindet also auch dann nicht, wenn alle hier vorhandenen Prädiktoren aus dem Kriterium partialisiert werden. Bei Math/Inf war dieser lineare Trend von einer leichten Tendenz überlagert, daß ein sehr langes Studium wieder mit eher besseren Zensuren zusammenhängt. Für Medizin ergab sich ein leichter kurvilinearer Zusammenhang: Studenten mit mittlerer Studiendauer erhielten die besseren Noten.

3. Das Merkmal "Prüfungsangst" hing mit den unter Auspartialisierung "aller Prädiktoren" gewonnenen Residuen nur bei den Angehörigen des Fachs Math/Inf (linear) zusammen: Studenten mit höherer Angst erhielten schlechtere Zensuren als aufgrund "aller Prädiktoren" vorhergesagt wurde.

4. "Das Verhältnis zu den Lehrern" hing bei den Medizinstudenten in schwer zu interpretierender Weise kurvilinear mit den Residuen, die sich nach Auspartialisierung der verschiedenen Prädiktorgruppen ergaben, zusammen. (Möglicherweise - diese

Vermutung ist jedoch empirisch nicht eindeutig zu sichern - sind die Examenszensuren am besten, wenn die Beziehungen zu den Lehrern entweder mittel bis schlecht oder besonders gut sind.)

5. Das Merkmal "soziale Schicht" zeigte keinerlei Effekte außer beim Studienfach Medizin. Hier ist der lineare Effekt, der durch die Einführung dieses Merkmals mit den "harten" Prädiktoren in die Regressionsgleichung auspartialisiert wurde, durch einen relativ starken, aber kaum interpretierbaren zickzackförmigen Resteffekt überlagert.

6. Medizinstudenten, die als Studienanfänger in mittlerem Ausmaß Kontakt zu Kommilitonen fanden, erreichten eher ein besseres Physikum, als aus der Kenntnis "aller" Prädiktoren erwartet wurde.

7. Ein wesentlicher Befund, weil dadurch Fragen des optimalen Analysemodells berührt werden, betrifft Streuungsänderungen der Residuen in Zusammenhang mit verschiedenen Merkmalen einschließlich der Schätzwerte. Diese Frage wird in ähnlicher, aber effizienterer Form in 5.4 weiterbehandelt. Hier zeigte sich bei den Studenten verschiedener Fächer mit standardisiertem Kriterium und bei den Studenten der BWL/VWL eine Tendenz zu kleineren Schätzfehlern bei allen Prädiktorgruppen mit schlechteren vorhergesagten Zensuren und schlechteren Abiturnoten. Bei Studierenden der Math/Inf waren die Schätzfehler bei allen Prädiktorgruppen am niedrigsten bei mittleren vorhergesagten Zensuren. Für Medizinstudenten wurde eine Tendenz zu kleineren Schätzfehlern bei allen Prädiktorgruppen nur in Zusammenhang mit schlechterer Selbstbeurteilung der Intelligenz gefunden. Die beobachteten Streuungsunterschiede der Residuen waren gering im Ausmaß - die Unterschiede betrugen maximal 20 % des größeren Wertes - überzeugten jedoch durch die Konsistenz ihres Auftretens bei den verschiedenen Prädiktorgruppen. Es bestehen also klare Hinweise auf Heteroskedastizitäten geringeren Ausmaßes.

5.4 Moderatoranalyse nach GHISELLI

Diese Methode ist die bisher erfolgreichste aller in größerem Umfang angewendeten Moderatortechniken (ZEDECK, 1971 a). Allerdings wurden praktische Erfolge häufig in Verbindung mit Methoden der Item-Analyse zur Entwicklung der Moderatorvariablen aus Fragebögen (bei GHISELLI Fragebögen mit Zwangwahlaufgaben) erzielt, während die Verwendung von bestehenden psychologischen Meßverfahren weniger leicht zur Identifikation von Moderatoren führt (GHISELLI, 1972).

GHISELLIs Methode ist einer Analyse von Residuen relativ ähnlich, und beide Methoden wurden auch gelegentlich miteinander verwechselt. Im Hauptunterschied zur Analyse der Residuen handelt es sich bei GHISELLI um Differenzen der standardisierten Schätz- und Kriteriumswerte, die nur absolut gewertet werden. Bei GHISELLIs Differenzwerten handelt es sich also um reine Maße der Schätzgenauigkeit ohne Berücksichtigung, ob Über- oder Unterschätzung vorliegt; Summen bzw. Durchschnitte dieser Differenzwerte sind in etwa Streuungen von Residuen vergleichbar (vgl. 5.3). Bei der Methode von GHISELLI werden nun Merkmale gesucht, die mit den genannten Differenzwerten korreliert sind, und die somit Personen mit guter Vorhersagbarkeit von Personen mit schlechter Vorhersagbarkeit zu trennen gestatten.

Bevor nun Strategien der Anwendung von Moderatoren überlegt werden, müssen die Daten auf das Vorhandensein solcher Effekte geprüft werden. Stellvertretend für alle recht ähnlich ausgefallenen Ergebnisse sollen die Befunde dargestellt werden, die sich aus der Berechnung der Schätzwerte aus den Abiturnoten ergaben. Eine Darstellung aller berechneten Korrelationen der Differenzwerte mit allen vorhandenen Variablen lohnt sich nicht, da die Koeffizienten insgesamt außerordentlich niedrig sind und nur in wenigen Fällen eine einigermaßen nennenswerte Höhe erreichen. Nachfolgend werden die auf dem 0,05 Niveau "signifikanten" Korrelationen mit den berechneten Zufallswahrscheinlichkeiten (jeweils eingeklammert) aufgezählt. Die Zufallswahrscheinlichkeiten können lediglich als beschreibende Kennwerte gelten, weil die absoluten Differenzwerte

ebenso wie die Residuen bereits mit wesentlich höheren Risiken verbunden sind als gewöhnliche Meßwerte.

Studierende in den 8 größten Fächern

(Kriterium fachweise standardisiert):

Abiturnote Deutsch	- 0,07	(0,003)
Abiturnote Biologie	- 0,06	(0,021)
Verhältnis zu den Lehrern	0,06	(0,041)
Schätzwert	- 0,06	(0,015)

BWL/VWL:

Abiturnote Deutsch	- 0,10	(0,022)
Abiturnote Mathematik	- 0,09	(0,041)
Mathematische Begabung	0,10	(0,04)
Schätzwert	- 0,11	(0,006)

Medizin:

Allgemeine Intelligenz	0,13	(0,029)

Math/Inf:

Abiturnote Physik	- 0,14	(0,027)
Vertrauenspersonen Geschwister	0,25	(0,001)

Für eine praktische Nutzung dürften die Koeffizienten vorwiegend zu gering sein. Sie bestätigen jedoch die bereits in 5.4 konstatierten Heteroskedastizitäten geringeren Ausmaßes, und sie unterstützen in der Tendenz die Vermutung von GHISELLI, daß sich die Individuen abweichend von der klassichen Theorie systematisch im Ausmaß der Fehlerstreuung unterscheiden.

Es folgen Bemerkungen zu den einzelnen der sich andeutenden Effekte:

1. Generell besteht eine Tendenz, daß die Vorhersagefehler mit schlechteren vorhergesagten Kriteriumswerten, mit schlechteren Abiturnoten und mit schwächerer Begabung geringer werden.

2. Bei den Studierenden der 8 größten Fächer insgesamt und bei den Studierenden der BWL/VWL hängen die Vorhersagefehler mit den Schätzwerten, mit Merkmalen, aus denen die Schätzwerte gebildet wurden und mit Merkmalen, die mit den Schätzwerten hoch korreliert sind, zusammen. Es handelt sich also um Heteroskedastizitäten, die mit den Prädiktoren selbst zusammenhängen; die Prädiktoren sind ihre eigenen Moderatoren.

3. Bei Medizin wird die Heteroskedastizität durch allgemeine Intelligenz, also ebenfalls ein Merkmal, das mit den wichtigsten Prädiktoren hoch korreliert ist, moderiert.

4. Das Studienfach Math/Inf zeigt ein von den anderen Fächern abweichendes Bild, bei allerdings großer Unsicherheit durch die Stichprobe von nur 246 Studierenden. Hier treten zwar als Moderatoren zwei Merkmale auf, die auch bei den harten Prädiktoren in der Regressionsgleichung vorhanden waren, jedoch stimmen die Richtungen der Moderatoreffekte im Verhältnis zu den Richtungen der Vorhersage in der Regressionsgleichung nicht überein:

   Mit schlechteren Zensuren im Vordiplom hängen die Eigenschaften "schlechtere Physikzensuren" und "Geschwister als Vertrauenspersonen" zusammen; geringere Schätzfehler dagegen wurden durch "schlechtere Physiknoten" und "Geschwister nicht als Vertrauenspersonen" vorhergesagt. Infolgedessen wurde die Heteroskedastizität auch nicht in der üblichen Weise vorhergesagt; in 5.3 zeigte sich eine kurvilineare Streuungsänderung der Residuen mit den Schätzwerten, die hier wegen der Verwendung von Produkt-Moment-Korrelationen nicht erfaßt werden konnte. Zu beachten ist bei diesen Moderatoreffekten, daß wegen der kleinen Stichprobe und wegen der extremen Verteilung des Merkmals "Vertrauenspersonen Geschwister" trotz ihrer Stärke eine praktische Anwendung mit zu großen Unsicherheiten verbunden wäre. Es müßte also erst die Bestätigung der Befunde an einer größeren Stichprobe, die auch eine Kreuzvalidierung zuließe, abgewartet werden.

5.5 Diskussion und Zusammenfassung

In diesem Kapitel wurde der Versuch unternommen, die Vorhersage des Studienerfolgs durch die Einführung nichtlinearer Effekte im Rahmen eines erweiterten linearen Modells zu verbessern. Dazu wurden folgende drei Ansätze erprobt:

1. Analyse von Produktvariablen innerhalb der üblichen Regressionsgleichung (Moderierte Regression nach SAUNDERS, 1956),
2. Suchen von kurvilinearen Zusammenhängen zwischen Residuen aus Regressionsgleichungen, die einen Teil der möglichen Prädiktoren enthalten, mit weiteren Prädiktoren ("piecewise regression") und
3. Aufspalten der Stichproben in Untergruppen mit unterschiedlicher Vorhersagbarkeit (Moderatoranalyse nach GHISELLI, 1960).

Die erzielten Effekte waren gering, und es besteht wohl nur geringe Aussicht - jedenfalls unter der dieser Arbeit zugrundeliegenden Auswahl von Variablen -, die Vorhersage des Studienerfolgs durch Erweiterungen des linearen Modells des hier erprobten Typs wesentlich zu verbessern. Vermutlich sollte ein solcher, neu zu unternehmender Versuch nicht nur von vorhandenen Merkmalen und Meßverfahren ausgehen, sondern es sollten Moderatoren gezielt, z.B. unter Einsatz von Techniken der Itemanalyse, zu entwickeln versucht werden.

Im einzelnen betrachtet blieb die Methode der moderierten Regression völlig erfolglos. Auch in anderen Untersuchungen war diese theoretisch elegante und effiziente Methode vermutlich wegen ihrer zu hohen Ansprüche an die Regelhaftigkeit der zu entdeckenden Strukturen in den Daten vorwiegend wenig erfolgreich (ZEDECK, 1971 a). Das Ergebnis bestätigte außerdem innerhalb der Befunde von 3.3 das Vorliegen reiner Haupteffekte im Sinn der Varianzanalyse.

Die Methode der "piecewise regression" erbrachte einige Aufschlüsse; jedoch waren keine - bis auf die im nächsten Absatz zu besprechende Heteroskedastizität - die Verbesserung der in 3.3 entwickelten Regressionsgleichungen direkt betreffenden Folgerungen daraus zu ziehen.

Bei der Moderatoranalyse nach GHISELLI wurden ebenfalls keine direkt in die Praxis umsetzbaren stärkeren Effekte erzielt. Jedoch ergaben sich hier vielleicht auch für spätere Arbeiten interessante Hinweise. Es zeigte sich, daß die Fehlerstreuungen in geringem Ausmaß aber in systematischer Weise mit bestimmten Variablen variierten; diese schwachen Moderatoreffekte erscheinen durch Übereinstimmungen zwischen den einzelnen Studienfächern und zwischen den Analysen unter Einsatz der verschiedenen Prädiktorgruppen als überzufällig sicher. Die Moderatoren stammten jeweils aus dem Bereich der ausgewählten Prädiktoren oder mit ihnen hoch korrelierender Variablen, d.h. die Prädiktoren sind vorwiegend ihre eigenen Moderatoren. Inhaltlich gesehen treten bei schlechteren vorhergesagten Zensuren geringere Schätzfehler auf. Dieses Ergebnis stimmt mit dem von FESTERLING (1971, nach TROST, 1975) berichteten Befund überein, daß der Zusammenhang zwischen Abiturnoten und Zensuren der ersten juristischen Staatsprüfung besonders eng im Bereich schlechter Zensuren ist. Die gefundenen Moderatoreffekte unterstützen den Vorschlag von GHISELLI, das klassiche Prädiktionsmodell mit dem Postulat der Homoskedastizität aufzugeben, und stattdessen nach Variablen zu suchen, die die Fehlerstreuung moderieren können. Vorzugsweise sollte dabei systematisch nach Moderatoren gesucht und sollten Moderatoren unter Einsatz von Techniken der Item-Analyse entwickelt werden.

Die Auswirkungen der geringen Grade von Heteroskedastizität auf die Gültigkeit des regressionsanalytischen Modells mit seiner Voraussetzung der Homoskedastizität erscheinen wenig einschneidend. Ohnehin bleibt die Erwartungstreue der Parameterschätzungen unberührt; lediglich die Effizienz, also die "Varianz" der Parameterschätzungen ist bei Heteroskedastizität herabgesetzt und die statistischen Tests verlieren damit im Prinzip ihre Gültigkeit (z.B. JOHNSTON, 1963). Nach Modellrechnungen von

JOHNSTON (1963, S. 215 ff) unter verschiedenen Graden der Heteroskedastizität ist jedoch von den hier gefundenen Ausmaßen nur ein minimaler und zu vernachlässigender Einfluß auf die Effizienz der Parameterschätzungen zu erwarten. Ansonsten wäre wohl der empfehlenswerteste Weg, Heteroskedastizitäten im Modell zu berücksichtigen (JAQUEZ et al., 1968), den "generalized least squares estimator" von AITKIN (nach JOHNSTON, 1963) zugrundezulegen. Voraussetzung dafür wären dann allerdings unabhängig gewonnene Angaben über Art und Ausmaß der Heteroskedastizität.

Auffällig ist inhaltlich gesehen, daß auch so stark "moderatorverdächtige" Merkmale wie Geschlecht und Sozialstatus nicht einmal eine Andeutung von Effekten erzielten.

Ein klares Ergebnis schließlich wurde mit der Tatsache gefunden, daß sich Angehörige verschiedener Studienfächer auch nicht nach fachweiser Standardisierung des Kriteriums und der Prädiktoren in günstiger Weise in ein und derselben Regressionsgleichung berücksichtigen lassen. Die einzelnen Fächer werden dabei in systematischer Weise über- und unterschätzt.

6. Die Kontrastgruppenanalyse (AID)

6.1 Komplexe Merkmalzusammenhänge

Bisher wurden in dieser Arbeit folgende, begrenzte Erweiterungen des linearen Modells durch nichtlineare Effekte erprobt: kurvilineare Beziehungen zwischen je zwei Merkmalen, kurvilineare Beziehungen zwischen Prädiktoren und Residuen, Interaktionen nach Art der Varianzanalyse (entsprechend der Moderatoranalyse von SAUNDERS, 1956) und Moderatoreffekte nach der Methode von GHISELLI (1960).

Wie nachfolgend ausgeführt, gibt es jedoch wesentlich komplexere Zusammenhänge mehrerer Merkmale, und es liegen verschiedene Verfahren vor, die geeignet sind, solche komplexen Zusammenhänge zu entdecken. Erste Hinweise und Behandlungsvorschläge stammen von MEEHL (1950); eine mathematische Fassung der vorgeschlagenen Verfahren besorgte HORST (1954). In Deutschland wurden solche Verfahren von LIENERT (1968) aufgegriffen und wesentlich weiter entwickelt; durch seine Arbeiten wurde ein größerer Kreis von Forschern mit diesen Verfahren bekanntgemacht, zu Anwendungen angeregt und überhaupt für Überlegungen über höherdimensionale Zusammenhänge aufgeschlossen.

Das "MEEHLsche Paradoxon" - ein konstruiertes Beispiel eines dreidimensionalen Zusammenhangs bei paarweiser Unabhängigkeit der beteiligten Variablen - wurde von LIENERT (KRAUTH und LIENERT, 1973 a) in Form der Abb. 6.1 dreidimensional dargestellt.

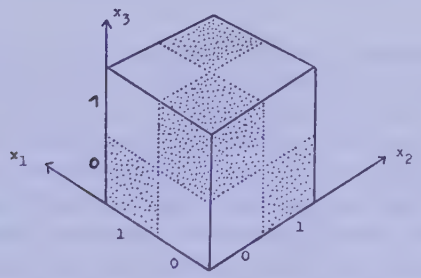


Abb. 6.1: (aus KRAUTH und LIENERT, 1973 a)
Das MEEHL'sche Paradoxon als Assoziation in einer dreidimensionalen Kontingenztafel mit x_1 und x_2 als Prädiktoren (binär verteilt) und x_3 als prädiziertem Kriterium (ebenfalls binär verteilt). Die 4 punktierten Oktanten sind mit Pbn besetzt, die 4 leeren Oktanten sind unbesetzt.

x_1 und x_2 sind als dichotome Prädiktoren, x_3 ist als dichotomes Kriterium aufzufassen. Die 4 punktierten Oktanten der dreidimensionalen Kontingenztafeln sind gleichmäßig mit Individuen besetzt, die 4 leeren Oktanten sind unbesetzt. Es zeigt sich - bei Projektion des Würfels jeweils auf eine Ebene -, daß keine paarweisen Zusammenhänge vorhanden sind. Dagegen läßt sich das Kriterium aus der gleichzeitigen Berücksichtigung beider Prädiktoren vollständig vorhersagen: Das Kriterium hat den Wert eins, wenn beide Prädiktoren den Wert eins oder null haben; es hat den Wert null, wenn die Prädiktoren die Wertkombination 1/0 oder 0/1 aufweisen. Von MEEHL wurde das Kriterium statt durch die einzelnen oder eine lineare Kombination der einzelnen Prädiktoren durch die Konfiguration der Prädiktoren in Form eines konfiguralen Validitätskoeffizienten, d.h. einer Korrelation der Konfigurationen mit dem Kriterium vorhergesagt. LIENERT dagegen schlägt ein Verfahren vor, das von einer multidimensionalen Kontingenztafel ausgeht und bei der Beurteilung der Häufigkeiten der Konfigurationen auf den Randhäufigkeiten der einzelnen Merkmale basiert und nicht nur die gegebenen Konfigurationen allein zugrunde legt.

Die von LIENERT entwickelten Verfahren sollen in Kap. 7 besprochen und erprobt werden; entscheidend ist hier die Tatsache, daß es höherdimensionale Wechselwirkungen geben kann, die bei niedriger dimensionaler Betrachtung oder bei der Analyse mit linearen Modellen unentdeckt bleiben. Vorwiegend wurden solche mehrdimensionalen Interaktionen in Kontingenztafeln z.B. allgemein von LANCASTER (1969) oder speziell für dreidimensionale Vierfeldtafeln von VICTOR (1972) behandelt. D.h. im Skalenniveau der beteiligten Merkmale wurde lediglich Kategorialniveau vorausgesetzt. Möglichkeiten, bestimmte, begrenzte Arten solcher Interaktionen auch bei quantitativen Skalen zu untersuchen, bieten die verschiedenen Moderatorkonzepte (vgl. Kap. 5). In diesem Kapitel soll mit der Kontrastgruppenanalyse (AID) ein Verfahren erprobt werden, das durch die Annahme eines höheren als nominalen Skalenniveaus eine effiziente Alternative

zur Regressionsanalyse darstellt, und das zusätzlich eine - wenn auch wiederum in gewisser Hinsicht begrenzte - Möglichkeit zur Entdeckung komplexer Merkmalszusammenhänge bietet.

6.2 Das Verfahren der Kontrastgruppenanalyse (AID)

Das hier Kontrastgruppenanalyse (nach ALLERBECK, 1972) oder AID genannte Verfahren ist zusätzlich unter verschiedenen Namen bekannt:

"Automatic interaction detector", abgekürzt "AID" (s.o.), durch SONQUIST und MORGAN (1964), sowie SONQUIST et al. (1973), die Protagonisten der Methode und Verfasser des vielseitigsten und bekanntesten Computerprogramms,

"Tree analysis" oder "Baum-Analyse" nach dem bei einer graphischen Darstellung der Ergebnisse entstehenden Bild,

"Auspik-Analyse" (abgekürzt nach "Automatischer Split zur optimalen Kombination erklärender Faktoren") von BIERVERT et al. (1969).

Gelegentlich wurden auch Bezeichnungen wie "multiple classification", "multilevel crosstabulations" und "configurational analysis" (nach ARMSTRONG et al., 1970) verwendet.

Die Methode, die in ihrer heutigen Form zuerst von MORGAN und SONQUIST (1963) beschrieben wurde, ist bereits - trotz ihres geringen Alters - zu einem häufig benutzten Verfahren vor allem in der empirischen Sozialforschung und in der Soziologie geworden. Sie geht von einer mindestens intervallskalierten abhängigen, und einer oder mehreren quantitativen oder qualitativen unabhängigen Variablen aus. Das Ziel des Verfahrens ist es, die Ausgangsstichprobe durch eine Folge von Zweiteilungen in mehrere Untergruppen aufzuspalten, die jeweils in bezug auf die abhängige Variable möglichst homogen sind. Jede Teilung einer Gruppe in zwei

Untergruppen ist so durch einen Prädiktor bestimmt, daß ein Teil seiner Kategorien die eine, und der Rest seiner Kategorien die andere Untergruppe definiert. Z.B. könnten einer Untergruppe die Personen mit besser als befriedigend in Englisch, der anderen Untergruppe die Personen mit befriedigend oder schlechter in Englisch angehören.

Die Prädiktoren werden je nachdem, ob sie qualitativ oder quantitativ sind, unterschiedlich behandelt. Bei einem "monotonen", quantitativen Prädiktor sind die Kategorien (siehe vorstehendes Beispiel) in einer bestimmten Reihenfolge angeordnet, so daß alle Kategorien des definierenden Prädiktors der einen Untergruppe größer bzw. kleiner als die der anderen Untergruppe sind. Bei einem "freien", qualitativen Prädiktor ist in seinen Kategorien keine bestimmte Reihenfolge enthalten, so daß hier die Kategorien im Prinzip frei zusammengestellt werden können, um die Untergruppen zu definieren. Das am häufigsten verwendete Programm von SONQUIST et al. (1973) ordnet die Kategorien eines freien Prädiktors nach ihren Mittelwerten in der abhängigen Variable und behandelt sodann den freien Prädiktor wie einen monotonen. Diese Behandlung eines freien Prädiktors erscheint als etwas "kühner" Schritt und ist zweifellos mit erhöhter statistischer Unsicherheit verbunden. Jedoch ist dieses Vorgehen unter bestimmten Voraussetzungen logisch durchaus zu rechtfertigen und für die Praxis sehr einfach zu handhaben. Das Problem verbleibt dann für die Interpretation.

Die Analyse beginnt mit einer Aufspaltung der Ausgangsstichprobe in zwei Untergruppen. Von allen möglichen Teilungen wird diejenige ausgewählt, die die durch beide Untergruppen erklärte Quadratsumme (bein SONQUIST et al., 1973) oder den t-Wert (bzw. den entsprechenden F-Wert) eines t-Tests zwischen den beiden Untergruppen maximiert. Dazu wird zunächst bei jedem einzelnen Prädiktor jene Stelle der Skala bestimmt, bei der eine Dichotomisierung den höchsten F-Wert bzw. die höchste Zwischen-Gruppen-Quadratsumme ergibt. Anschließend wird der nach dem gleichen Gesichtspunkt effektivste Prädiktor von allen zur Aufspaltung der Ausgangsstichprobe in die beiden Untergruppen benutzt. Nach-

folgend wird jede der dabei entstandenen Untergruppen erneut aufgespalten, usw., bis ein Abbruchkriterium - meist die Unterschreitung einer minimalen Gruppengröße oder zu geringe Unterschiedlichkeit der Gruppen - erreicht ist. Der Nachteil des Informationsverlustes bei der Dichotomisierung quantitativer Skalen wird zum Teil dadurch ausgeglichen, daß die restlichen Teile der Skala eines Prädiktors, der schon einmal als Basis für eine Teilung diente, weiterhin in den entsprechenden Untergruppen als potentielle Prädiktoren geführt werden. Die Teile eines bereits einmal zur Aufspaltung einer Gruppe benutzten Prädiktors können also erneut - falls ein entsprechend hoher Erklärungswert vorliegt - zur Basis von Aufteilungen gemacht werden, so daß theoretisch die ganze Spannweite eines quantitativen Prädiktors zur Vorhersage genutzt werden kann.

Das Ergebnis des Aufspaltungsverfahrens wird in einem sogenannten "Baumdiagramm" wie z.B. Abb. 6.2 dargestellt. Die in den einzelnen Stadien der Analyse durchgeführten Zweiteilungen bilden meist eine Hierarchie von den effektiveren (links in der Abbildung 6.2) zu den weniger effektiven (rechts) Zweiteilungen, sofern keine höherdimensionalen Beziehungen zwischen den Merkmalen vorliegen. Komplexere Wechselwirkungen zwischen den Merkmalen würden sich in entsprechenden Abfolgen der Teilungen zeigen und könnten zu besonders hohen F-Werten in späteren Teilungsstadien führen. Voraussetzung zur Entdeckung von höherdimensionalen Wechselwirkungen ist jedoch stets, daß dabei beteiligte Merkmale zunächst zur Basis von einfachen Teilungen gemacht werden können.

Die sich rechts an den Endpunkten der Verzweigungslinien (z.B. der Abb. 6.2) befindenden Untergruppen bilden eine vollständige Aufteilung der Stichprobe; für jede dieser Untergruppen werden unterschiedliche Kriteriumswerte vorhergesagt. Ein mögliches Verfahren zur Beschreibung dieser Aufspaltung der Stichprobe - nicht aber zur inferenzstatistischen Beurteilung wie BIERVERT et al. (1969) annehmen[1] - ist die einfaktorielle Varianzanalyse.

---

1) Ebensowenig ist natürlich der t-Test ein geeignetes Verfahren zur inferenzstatistischen Beurteilung der Spaltung einzelner Gruppen in Untergruppen.

Für den praktischen Vergleich der Ergebnisse einer Kontrastgruppen - und einer Regressionsanalyse ist die direkte Entsprechung des Eta-Quadrat-Koeffizienten der Varianzanalyse (z.B. HAYS, 1963) mit der quadrierten multiplen Korrelation (COHEN, 1969) von Vorteil. Problematischer ist allerdings der statische Vergleich der Ergebnisse beider Verfahren. Da bis vor kurzem überhaupt keine statistischen Tests für AID bekannt waren, war man ausschließlich auf Kreuzvalidierungen angewiesen. Neuerdings wurden allerdings von KASS (1975) und SCOTT und KNOTT (1976) Ansätze zu statistischen Modellen entwickelt, die AID über ein wie bisher rein beschreibendes Verfahren hinaus erweitern. Die entwickelten näherungsweisen Tests sollen in den durchzuführenden Analysen benutzt werden und für eine bessere Beurteilung der Ergebnisse Sorge tragen.

Ein entscheidender Nachteil der Kontrastgruppenanalyse ist die - bisher durch allgemeine, praktische Erfahrungen und Überlegungen begründete - Notwendigkeit, sehr umfangreiche Stichproben zu verwenden: z.B. vermuten ARMSTRONG et al. (1970), daß bei 5 Prädiktoren, von denen jeder zur Grundlage einer Zweiteilung gemacht werden soll, eine Stichprobe von "wesentlich mehr als 1000 Personen" nötig ist, d.h. die Stichprobe muß "wesentlich mehr als" 5 mal so groß wie bei einer entsprechenden Regressionsanalyse sein.

Vorteilhaft sind dagegen die feinere Anpassung von AID an die Struktur der Daten und die Möglichkeit, unter der Voraussetzung der Beteiligung relevanter Prädiktoren an vorangegangenen Teilungen komplexe Berechnungen zwischen den Variablen à la MEEHL zu entdecken. Noch scheinen allerdings keine praktischen Erfahrungen über die Fähigkeit von AID zur Identifikation solcher Beziehungen vorzuliegen, und sie sind auch nicht das erklärte Ziel der vorhandenen Untersuchungen, die sich eher um eine effiziente Alternative zur Regressionsanalyse im allgemeinen Sinn bemühen.

Ein weiterer Vorzug der Kontrastgruppenanalyse ist nicht zuletzt die anschauliche Darstellbarkeit der Ergebnisse in einem Baumdiagramm, das auch statistisch weniger Erfahrenen nur geringe Verständnisschwierigkeiten bereitet. Sicherlich ist diese Eigenschaft, neben der Eignung des Verfahrens, von fast jeder Art von Daten in irgendeiner Hinsicht interessante Ergebnisse zu liefern, ein wichtiger Grund für die Beliebtheit von AID in der angewandten Sozialforschung.

6.3 Ergebnisse

Als Rechenprogramm wurde zunächst das bekannte und vielseitige AID 3 aus dem Programmpaket OSIRIS (ICPR, 1971) benutzt. Nachdem nicht behebbare Schwierigkeiten aufgetreten waren, wurden nur abgesicherte Teilergebnisse aus diesem Programm verwendet, und es wurde in der Hauptsache auf das Programm Auspik (BIERVERT et al., 1969) in einer von Dipl.Math. KÖGEL adaptierten Fassung zurückgegriffen. Als Abbruchkriterium wurde eine Gruppengröße von 20 gesetzt. Die endgültigen Lösungen ergaben sich nach Anwendung der näherungsweise gültigen Signifikanztests von KASS (1975) und SCOTT und KNOTT (1976) bei einem Irrtumsrisiko von 0,05 für einzelne Studienfächer und von 0,01 für die Gesamtgruppe. Die genannten Tests gestatten die statistische Beurteilung jeweils einer einzelnen Zweiteilung. Eine Adjustierung des Alpha-Risikos für die Tatsache der schrittweisen Einführung von jeweils einem aus einer Gruppe von möglichen Prädiktoren ausgewählten Prädiktor wurde in Parallelität zum Vorgehen bei der Regressionsanalyse nicht vorgenommen. Ebenso wie bei der Regressionsanalyse sind also die einzelnen Effekte als heuristisch zu interpretieren.

Besondere Probleme ergaben sich bei dem Einschluß des qualitativen Merkmals "Vertrauenspersonen" im Rahmen der harten Prädiktoren:

1. Innerhalb jeder der analysierten Stichproben wurden die Kategorien des Prädiktors "Vertrauenspersonen" nach ihren Mittelwerten in der abhängigen Variable geordnet, so daß das Merkmal wie ein monotoner Prädiktor behandelt werden konnte. Die Reihenfolge der Kategorien differierte deutlich bei den verschiedenen Stichproben; insofern ergab sich keine die Logik des Verfahrens unterstützende Ausgangsbasis.

2. Bei der Durchführung der Analysen wurde der Prädiktor "Vertrauenspersonen" gemessen an den Signifikanztests von KASS (1975) und SCOTT und KNOTT (1976) im Verhältnis zu den anderen (wirklich monotonen) Prädiktoren systematisch bevorzugt: Die Auswahl des bei jeder Zweiteilung besten Prädiktors wurde vom Programm nach den F-Werten ohne Rücksicht darauf, ob es sich um einen freien oder monotonen Prädiktor handelte, vorgenommen. Da aber die Ergebnisse der angemessenen, genannten Signifikanztests bei freien Prädiktoren stärker von den Ergebnissen der einfachen F-Tests abweichen als bei monotonen Prädiktoren, werden freie Prädiktoren aufgrund der Auswahl des Programms nach einfachen F-Tests in unangemessener Weise bevorzugt.

Weil eine saubere Lösung dieses Problems - nämlich die Prädiktorenauswahl innerhalb des Programms nach den angemessenen statistischen Kriterien - aus technischen Gründen nicht zu leisten war, wurde ein gröberes Verfahren gewählt: Wenn in einer Analyse das Merkmal "Vertrauenspersonen" zwar als Prädiktor ausgewählt wurde, aber in der anschließenden statistischen Prüfung keine Signifikanz erreichte, wurde die Analyse ohne dieses Merkmal wiederholt.[1)]

Aufgrund des vorstehend Gesagten wird klar, daß das beschriebene Vorgehen nicht optimal sein kann, und daß erst durch ein neu zu schreibendes oder modifiziertes Programm, das die genannten statistischen Tests direkt nutzen würde, ein angemesseneres Vorgehen

---

1) Der im Prinzip ebenfalls mögliche Weg, den qualitativen Prädiktor Vertrauenspersonen in Dummyvariablen zu zerlegen, wurde auch versucht. Es ergaben sich jedoch keine abweichenden Ergebnisse dabei.

möglich wäre. Der Effekt der endgültigen Aufspaltung der Stichprobe in Untergruppen auf die Vorhersage des Kriteriums wurde jeweils mit Hilfe einer einfaktoriellen Varianzanalyse untersucht und in einem Eta-Quadrat-Koeffizienten ausgedrückt (HAYS, 1963).

Es wurden alle Studienfächer einzeln, sowie die Gesamtgruppe von 8 Studienfächern mit fachweise standardisiertem Kriterium analysiert. Dabei fanden nur die "harten" Prädiktoren Verwendung, da bei den "weichen" Prädiktoren die Zahl der Effekte im Verhältnis zum Stichprobenumfang mit Sicherheit zu groß geworden wäre. Außerdem sind bei der Verwendung von Prädiktoren aus unterschiedlichen Bereichen - hier Abiturnoten und andere persönliche Merkmale - am ehesten Wechselwirkungen zu erreichen (DIGMAN, 1966). Die Ergebnisse sind in den Abbildungen 6.2 bis 6.5 repräsentiert. Die Stichprobenumfänge der kleineren Studienfächer erwiesen sich als zu gering; deshalb wurde nur das Studienfach Math/Inf mit einem für die noch kleineren Fächer typischen Befund dargestellt. Die Bedeutung der Signifikanztests zeigte sich z.B. bei den kleineren Stichproben, wenn zunächst durchaus beträchtlich erscheinende Effekte auftraten, von denen nach Anwendung der statistischen Tests nur wenig übrig blieb. Zusätzlich zu den Abbildungen 6.2 - 6.5 wird zur Illustration eines möglichen Ergebnisses unter Verwendung der Abiturnoten als Prädiktoren in Abb. 6.6 die Gesamtgruppe von 8 Studienfächern dargestellt.

Die Ausgangsstichprobe ist jeweils im Kästchen ganz links mit Stichprobenumfang und Mittelwert des Kriteriums angegeben. Die anderen Kästchen bezeichnen die weiteren, sich entlang den Verzweigungslinien ergebenden Gruppen mit den endgültigen, die gesamte Stichprobe repräsentierenden Gruppen an den Enden der Verzweigungslinen. Jede Gruppe wird mit dem Teilungskriterium auf dem ausgewählten Prädiktor, ihrem Mittelwert in der abhängigen Variable und ihrem Umfang (angegeben mit n) bezeichnet. So werden z.B. die in Abb. 6.3 bei der ersten Teilung erzeugten Untergruppen durch 176 Studierende mit besser als befriedigend in Mathematik und einem Mittelwert im Kriterium von 2,44, sowie durch 503 Studierende mit schlechter als gut in Mathematik und einem Mittelwert im Kriterium von 2,85 gebildet.

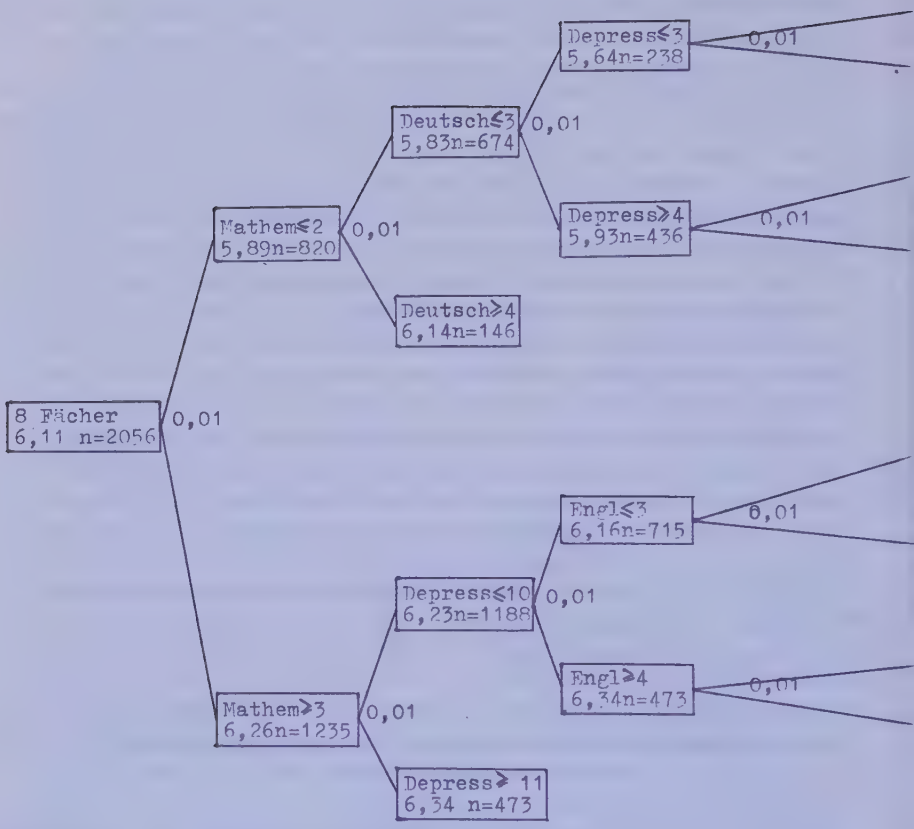


Abb. 6.2 (erste Hälfte): Kontrastgruppenanalyse der 8 größten Studienfächer (Kriterium fachweise standardisiert) unter Verwendung der harten Prädiktoren

Eta-Quadrat = 0,105

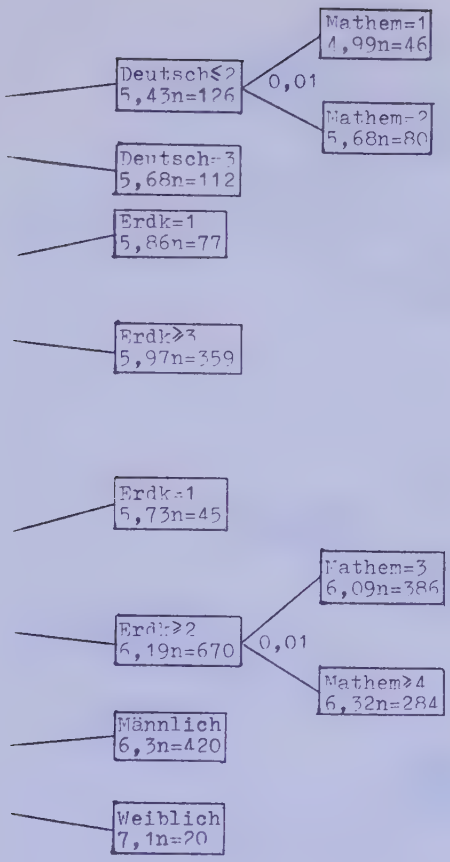


Abb. 6.2 (zweite Hälfte)

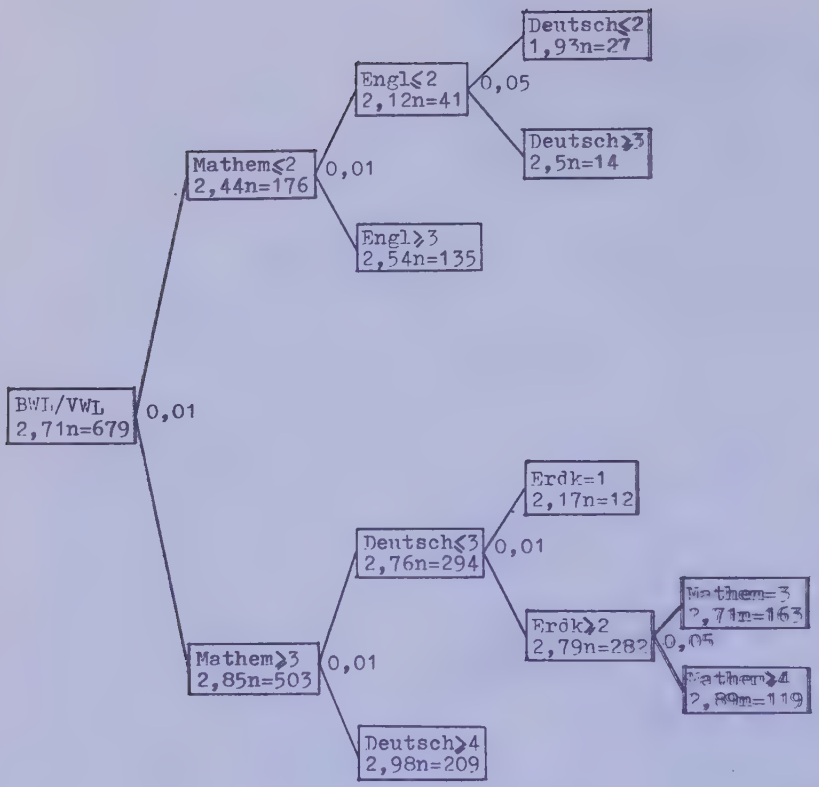


Abb. 6.3: Kontrastgruppenanalyse des Studienfachs BWL/VWL unter Verwendung der harten Prädiktoren

Eta-Quadrat = 0,16

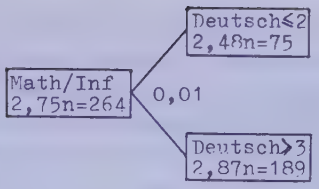


Abb. 6.5: Kontrastgruppenanalyse des Studienfachs Math/Inf unter Verwendung der harten Prädiktoren

Eta-Quadrat = 0,053

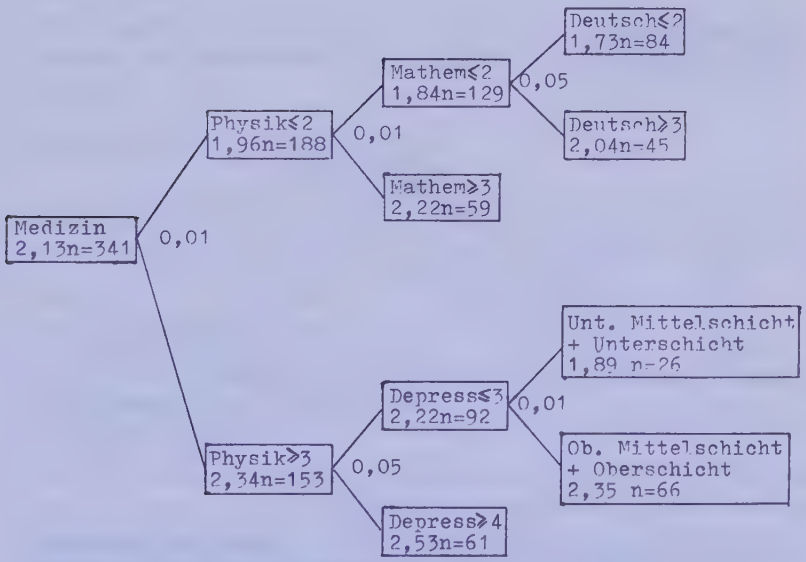


Abb. 6.4: Kontrastgruppenanalyse des Studienfachs Medizin unter Verwendung der harten Prädiktoren

Eta-Quadrat = 0,165

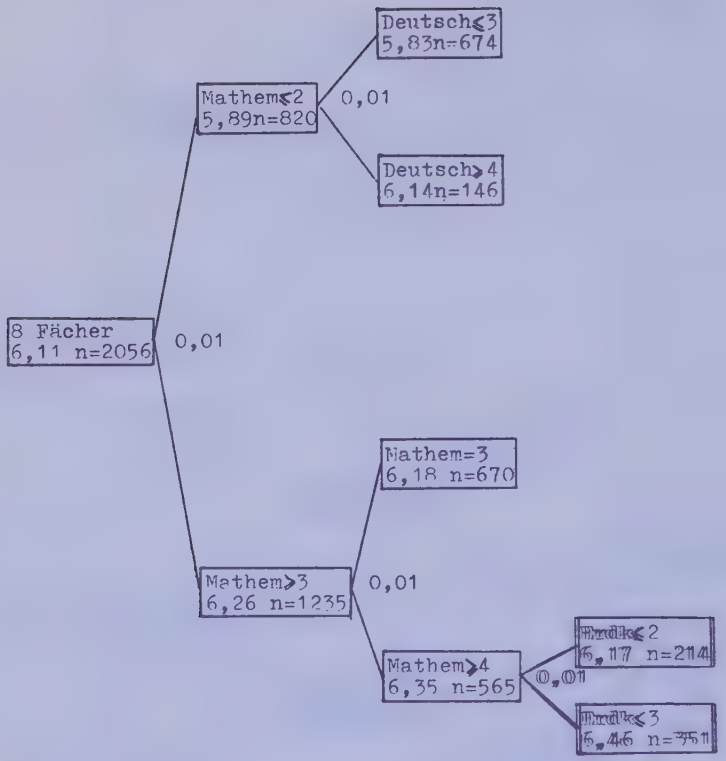


Abb. 6.6: Kontrastgruppenanalyse der 8 größten Studienfächer (Kriterium fachweise standardisiert) unter Verwendung der Abiturnoten als Prädiktoren

Eta-Quadrat = 0,071

Die Mittelwerte des Kriteriums sind in Abb. 6.2 und 6.6 bei den Gesamtgruppen in ihrer absoluten Höhe nicht interpretierbar; das Kriterium war zunächst - wie in 3.3 beschrieben - fachweise standardisiert worden und mußte hier aus programmtechnischen Gründen weiter transformiert werden, wobei jeder anschauliche Bezug der verwendeten Skala verlorenging.

In den Winkeln der Verzweigungslinien sind jeweils die überschrittenen Irrtumswahrscheinlichkeiten nach den Tests von KASS (1975) und SCOTT et al. (1976) angegeben.

6.4 Diskussion

Zur Gegenüberstellung der jeweils durch die Kontrastgruppenanalyse und Regressionsanalyse erklärten Varianzteile siehe Tabelle 6.1.

Tabelle 6.1

Prädiktoren	Stichprobe	R-Quadrat	Eta-Quadrat
harte Prädiktoren	8 Studienfächer	0,10	0,105
	BWL/VWL	0,123	0,16
	Medizin	0,123	0,165
	Math/Inf	0,12	0,053
Abiturnoten	8 Studienfächer	0,10	0,071

Zu beachten ist folgender Unterschied beim schrittweisen Aufbau von R-Quadrat und Eta-Quadrat: R-Quadrat wächst praktisch nicht über eine bestimmte Grenze hinaus, auch wenn viele weitere Prädiktoren in die Gleichung einbezogen werden. Eta-Quadrat wächst dagegen immer weiter, je mehr Teilungen durchgeführt, d.h. je mehr Prädiktoren genutzt werden. Offensichtlich wird die Zufallsvarianz durch AID sehr viel besser genutzt als bei der Regressionsanalyse. Diese Erklärung von Zufallsvarianz sollte nun durch die genannten statistischen Tests eingeschränkt werden. Allerdings war die Kombination von Test und Programm nicht

optimal (vgl. 6.3) und die Ergebnisse wurden deshalb durch die statistischen Tests stärker eingeschränkt als notwendig war. In einem günstigeren Programm, das die Signifikanztests bereits während der Durchführung der Analyse einsetzt, könnte also vermutlich ein grösserer Anteil der Kriteriumsvarianz erklärt werden.

Darüber hinaus ist bei dem durch AID erzeugten Eta-Quadrat keine Stichprobenverteilung bekannt, wie es z.B. für die multiple Korrelation der Fall ist. Bezüglich einer Abschätzung der zu erwartenden Schrumpfung ist man also weiterhin auf Kreuzvalidierungen angewiesen (vgl. 6.5).

In der Tab. 6.1 zeigen sich keine konsistenten Unterschiede zwischen den Ergebnissen beider Verfahren: Das Studienfach Math/Inf zeigt eindeutig den Effekt der für AID zu kleinen Stichprobe und ist insofern nicht weiter zu werten. Bei der Gesamtgruppe von 8 Fächern liegt Eta-Quadrat niedriger als R-Quadrat; hier ist die in 5.3 bestätigte Inhomogenität der Gesamtgruppe trotz Standardisierung zu berücksichtigen, die sich in schwer zu beurteilender Weise besonders auf den Aufspaltungsprozeß bei AID ausgewirkt haben könnte. Bei den Studienfächern BWL/VWL und Medizin sind am ehesten die Voraussetzungen für einen fundierten Vergleich beider Verfahren gegeben; hier erweist sich AID als überlegen.

Eine in der Zielsetzung vergleichbare Untersuchung von ARMSTRONG et al. (1970), die den Umsatz von ca. 6000 Tankstellen mit Hilfe von AID und multipler Regressionsanalyse vorherzusagen versuchte, erwies die Gleichwertigkeit beider Verfahren in bezug auf die erklärte Kriteriumsvarianz; jedoch unterlagen die Ergebnisse von AID einer geringeren Schrumpfung bei Kreuzvalidierung. Die Schwierigkeit, diese Ergebnisse mit den vorliegenden zu vergleichen, liegt außer in Art und Menge der Daten vor allem darin, daß von ARMSTRONG et al. keine Signifikanztests durchgeführt werden konnten: Möglicherweise wurde der Spaltungsprozeß bei AID von ARMSTRONG et al. früher abgebrochen, als es nach den Tests von KASS (1975) und SCOTT et al. (1976) nahegelegt worden wäre. Dieses "vorsichtige" Vorgehen könnte dann entsprechend zu einem nicht so sehr hohen

Anteil an erklärter Varianz und einer relativ geringen Schrumpfung geführt haben. Relativ einleuchtend ist jedoch die Folgerung von ARMSTRONG et al., daß bei der Anwendung auf seine Daten und ihnen ähnliche Datensätze AID das zu bevorzugende Verfahren darstellt.

Die Interpretation der einzelnen, in den Abbildungen 6.2 bis 6.6 enthaltenen Effekte wird ebenso wie und vielleicht sogar noch mehr als bei der multiplen Regressionsanalyse durch die hohen Multikollinearitäten der Prädiktoren erschwert. Wenn beispielsweise in Abb. 6.3 die Teilung in der ersten Ebene analog der Regressionsanalyse nach der Mathematikzensur erfolgt, und die Teilungen der zweiten Ebene nicht nur nach Deutsch wie bei der Regressionsanalyse, sondern nach Deutsch und Englisch erfolgen, liegt die Vermutung nahe, daß die Entscheidung zwischen Deutsch und Englisch in der zweiten Teilungsebene von minimalen und damit eher zufälligen Unterschieden abhängt. D.h. man muß keine wesentlich von den Ergebnissen der Regressionsanalyse abweichenden Befunde annehmen. Diese Vermutung einer schon von der Regressionsanalyse her bekannten hohen Kollinearität der Prädiktoren bei AID konnte durch die Kontrolle der bei jeder Teilung nicht ausgewählten Prädiktoren bestätigt werden. Wegen des hohen Aufwandes wurden diese Kontrollen nur stichprobenweise an ausgewählten Stellen der Analysen wie z.B. dem eben aufgeführten Beispiel durchgeführt. Eine klare und eindeutige Identifikation von Wechselwirkungen leichteren Grades scheitert also an der in den Daten enthaltenen Kollinearität; man muß die Fähigkeit von AID als "automatischem Interaktionsdetektor" wohl grundsätzlich durch jegliche Ausprägung von Multikollinearität eingeschränkt sehen. Komplexe Wechselwirkungen stärkeren Grades, die sich z.B. durch das Auftreten von nichtordinalen Wechselwirkungen oder von hervorstechend hohen F-Werten in späteren Teilungsebenen zeigen könnten, wurden in den Abbildungen 6.2 - 6.6 nicht entdeckt.

Die Auswahl der von AID benutzten Prädiktoren ist den Ergebnissen der Regressionsanalysen sehr ähnlich: Vorwiegend wurden (bei den harten Prädiktoren) Abiturnoten ausgewählt; gelegentlich treten

andere Merkmale auf, die - wie bei Medizin - auch bei den Regressionsanalysen ausgewählt worden waren oder - wie bei der Gesamtgruppe - bei den Regressionsanalysen knapp unter der Signifikanzschwelle gelegen hatten.

6.5 Kreuzvalidierung der Ergebnisse von multiplen Regressionsanalysen und von Kontrastgruppenanalysen

Der Hauptzweck der hier durchgeführten Berechnungen war der Vergleich der Ergebnisse von Kontrastgruppen - und Regressionsanalysen. Am interessantesten wäre dabei die Untersuchung der harten Prädiktoren gewesen: Jedoch hatten z.B. die Erfahrungen von 6.3 gezeigt, daß bei dieser Zahl von Prädiktoren eine Untersuchung von kleineren Stichproben als 300, die ja bei Halbierung von Stichproben zur Kreuzvalidierung besonders schnell unterschritten wird, kaum noch sinnvoll ist. So wurden unter der Verwendung der harten Prädiktoren nur die Gesamtgruppe von 8 Fächern und das Studienfach BWL/VWL, und unter Verwendung der Abiturnoten als Prädiktoren die Gesamtgruppe und die Studienfächer BWL/VWL, Medizin und Math/Inf einer Kreuzvalidierung bezüglich der Ergebnisse von Kontrastgruppen- und Regressionsanalyse unterzogen.

Die genannten Stichproben wurden mit Hilfe von Zufallszahlen halbiert. Das in jeweils einer Hälfte der Stichprobe entwickelte Vorhersageverfahren wurde in beiden Stichprobenhälften in Form einer Korrelation der Schätz- mit den Kriteriumswerten bei der Regressionsanalyse und in Form eines Eta-Quadrat-Koeffizienten von einfaktoriellen Varianzanalysen bei AID ausgewertet. Die Ergebnisse sind in Tab. 6.2 enthalten. Die Schrumpfung der multiplen Korrelationen ist relativ gering; teilweise ergibt sich sogar ein geringer - zufallsbedingter - Anstieg. Die Schrumpfung liegt bei den Kontrastgruppenanalysen durchweg höher, ist aber ebenfalls nicht übermäßig einschneidend oder gar jeden Effekt vernichtend. Der Befund von ARMSTRONG et al. (1970), daß bei AID

Tabelle 6.2: Kreuzvalidierung der Ergebnisse von multiplen Regressionsanalysen und Kontrastgruppenanlysen

		R-Quadrat		Eta-Quadrat	
Prädiktoren	Stichprobe	Analysestichprobenhälfte	Validierungsstichprobenhälfte	Analysestichprobenhälfte	Validierungsstichprobenhälfte
harte Prädiktoren	8 Fächer (Kriterium standardisiert)	0,09 0,08	0,078 0,086	0,093 0,091	0,074 0,079
	BWL/VWL	0,084 0,09	0,084 0,084	0,133 0,126	0,117 0,112
Abiturnoten	8 Fächer (Kriterium standardisiert)	0,09 0,08	0,078 0,086	0,074 0,081	0,057 0,052
	BWL/VWL	0,084 0,09	0,084 0,084	0,065 0,12	0,056 0,103
	Medizin	0,09 0,09	0,09 0,098	0,122 0,108	0,085 0,091
	Math/Inf	0,09 0,09	0,069 0,073	0,044 0,084	0 0,039

eine geringere Schrumpfung als bei der multiplen Regressionsanalyse zu beobachten ist, wird nicht bestätigt. Vermutlich ist diese Diskrepanz darauf zurückzuführen, daß von ARMSTRONG et al. (1970) im Vergleich mit der vorliegenden Untersuchung stärker konservativ vorgegangen wurde, so daß relativ weniger Kriteriumsvarianz erklärt wurde, die dann aber auch einer geringeren Schrumpfung unterlag.

Insgesamt läßt sich feststellen, daß die Kontrastgruppenanalyse in Bezug auf ihre Effektivität als Alternative zur Regressionsanalyse durchaus interessant bleibt.

6.6 Die Analyse von Residuen mit der Kontrastgruppenanalyse

Es wurde versucht, die Vorteile der Regressionsanalyse - als einem bewährten und effizienten Verfahren - sowie die Vorteile der Kontrastgruppenanalyse mit einer besseren Berücksichtigung nichtlinearer Effekte zu verbinden. Dazu wurden die linearen, mit Hilfe der Regressionsanalysen bestimmten, Effekte der "harten" Prädiktoren aus dem Kriterium partialisiert, und die berechneten Residuen wurden einer Kontrastgruppenanalyse mit den "harten" Prädiktoren unterzogen. Es sollte dadurch geprüft werden, ob sich nach dem Ausschalten der linearen Effekte weitere nichtlineare Effekte mit Hilfe von AID feststellen ließen. Das Verfahren bei der Durchführung von AID entsprach dem in 6.3 beschriebenen, und es wurden die gleichen Stichproben wie in Abb. 6.2 - 6.6 verwendet.

Die Ergebnisse waren negativ. Das Programm erzeugte zwar Teilungen in 4 - 6 Ebenen, d.h. aufgrund der Komplexität der Gruppenbildung wären durchaus Interaktionen höherer Ordnung möglich gewesen. Jedoch war keine der mit einer Teilung - auch nicht in einem späteren Teilungsstadium - verbundenen Irrtumswahrscheinlichkeiten kleiner als 0,05. Die erzielten nichtsignifikanten Tendenzen waren teilweise mit Tendenzen bei den Regressionsanalysen identisch.

Zusammenfassend soll für dieses Kapitel folgendes festgestellt werden: Es ließ sich keine klare Überlegenheit eines der beiden Verfahren Kontrastgruppen- und multiple Regressionsanalyse in bezug auf ihre Effizienz bei Kreuzvalidierung feststellen. Vermutlich waren die innerhalb der vorliegenden Daten analysierbaren Stichproben zu gering im Umfang, um eine solche Entscheidung fällen zu können. Bei der Anwendung von AID auf das Kriterium oder das von linearen Effekten bereinigte Kriterium konnten keine Hinweise auf höherdimensionale Wechselwirkungen stärkeren Ausmaßes gewonnen werden. Insgesamt bleibt eine Anwendung von AID auf ähnliche Daten interessant und wünschenswert[1)], da eine der Regressionsanalyse vergleichbare Effizienz zu erwarten ist, und weil außerdem eine bessere Anpassung an die Struktur solcher Daten möglich ist, die nicht genau den Voraussetzungen der multiplen Regressionsanalyse entsprechen.

7. Die Konfigurationsfrequenzanalyse (KFA)

Bei allen bisher verwendeten Verfahren waren die Möglichkeiten, völlig adimensionale Effekte und Wechselwirkungen höheren Grades zwischen den beteiligten Variablen zu entdecken, durch bestimmte Bedingungen eingeschränkt. Ein möglicher Weg, sämtliche Beschränkungen in dieser Hinsicht auszuschalten, besteht darin, alle möglichen Konfigurationen, die durch die Ausprägungen der Prädiktorvariablen gebildet werden können, einzeln zu berücksichtigen und für die Vorhersage zu nutzen. Dabei muß allerdings der quantitative Gesichtspunkt bei den entsprechenden Prädiktoren verlorengehen, weil bei der Bildung aller möglichen - formal gleichberechtigten - Konfigurationen keine Regeln verfolgt werden können, die eine Ordnung der Konfigurationen nach einem quantitativen Gesichtspunkt erlauben würden. Am weitesten in Richtung auf die Berücksichtigung konfi-

---

1) Wünschenswert wäre vor allem auch, Programme zu erstellen, die die Signifikanztests von KASS (1975) und SCOTT und KNOTT (1976) unmittelbar bei der Durchführung der Analyse berücksichtigen.

gurationaler Effekte ohne Aufgabe des quantitativen Gesichtspunktes und den damit verbundenen Effizienzverlust ging die im vorigen Kapitel besprochene Kontrastgruppenanalyse. Allerdings ist die Voraussetzung dafür, mit Hilfe dieses Verfahrens höherdimensionale Wechselwirkungen zu entdecken, daß diese Beziehungen durch Wechselwirkungen von jeweils niedrigerer Dimension überlagert sein müssen. Immerhin ist dies nach KRAUTH und LIENERT (1973 a, S. 24) der weitaus häufigste Fall.

Bei den Methoden, die Konfigurationen direkt zur Vorhersage eines quantitativen Kriteriums zu nutzen, kann man vor allem zwei Varianten unterscheiden: Die eine Methode ordnet jede Vp der ihr entsprechenden oder am besten passenden Konfiguration zu und verwendet diese eine Angabe über jede Vp als Basis zur Vorhersage (LUBIN und OSBORN, 1957, FOREHAND und MCQUITTY, 1959). Die zweite Variante, angeregt durch den bereits in 6.1 erwähnten Artikel von HORST (1954), weist jeder Vp eine Reihe von Meßwerten zu. Dies sind zusätzlich zu den Werten in den Ausgangsprädiktoren die Ausprägungen in den verschiedenen möglichen Zweierkonfigurationen, Dreierkonfigurationen, usw. Alle diese Meßwerte - eventuell nach einer Auslese der wichtigsten der möglichen Konfigurationen z.B. nach ihrer Reliabilität - bilden dann die Basis zur Vorhersage des Kriteriums etwa in Form einer multiplen Korrelation. Beide Varianten waren in der Praxis wenig erfolgreich; spätestens bei der Kreuzvalidierung fiel die Validität auf einen durchschnittlichen, auch ohne Berücksichtigung von Konfigurationen zu erreichenden Wert zurück (z.B. LUNNEBORG und LUNNEBORG, 1967).

Ein anderer Weg, konfigurationale Information zu nutzen, wurde von LIENERT (1968 und viele weitere Arbeiten, zusammenfassend KRAUTH und LIENERT, 1973 a), vor allem in Deutschland[1)] bekannt gemacht. Das Konfigurationsfrequenzanalyse oder abgekürzt KFA genannte Verfahren und seine Varianten sind Gegenstand des vorliegenden Kapitels.

---

1) Eine englischsprachige Vorstellung des Verfahrens erfolgte durch LIENERT und KRAUTH (1975).

7.1 Konfigurationsfrequenzanalyse (KFA) und Interaktionsstrukturanalyse

Das Verfahren der KFA, das hinreichend bekannt sein dürfte, um auf eine Beschreibung verzichten zu können, wurde zunächst und vor allem zur Findung von Typen eingesetzt. Ein wichtiger Vorteil dieser parameterfreien Methode ist die Berechnung der erwarteten Frequenzen der Konfigurationen aus den Häufigkeiten, mit denen die Realisierungen der einzelnen Variablen auftreten. Dabei wird das Modell einer mehrdimensionalen Kontingenztafel zugrundegelegt. Die Beurteilung der Häufigkeit der einzelnen Konfigurationen erfolgt also an Hand eines komplexen Modells, dessen Nullhypothese Unabhängigkeit aller beteiligten Merkmale annimmt. Dabei werden beliebig komplexe Beziehungen zwischen den Merkmalen berücksichtigt, und die einzelnen mehrdimensionalen Interaktionen können in der vorgeschlagenen Definition mit Hilfe der Assoziationsstrukturanalyse identifiziert werden.

Einschränkend auf die Anwendbarkeit des Verfahrens wirkt sich das ungünstige Verhältnis des erforderlichen Stichprobenumfangs zu Anzahl der beteiligten Merkmale und Zahl ihrer Kategorien aus, d.h. man muß entweder sehr umfangreiche Stichproben oder aber recht gut begründete Hypothesen bezüglich der relevanten Merkmale zur Verfügung haben. Der zweite genannte Fall dürfte am ehesten im Bereich der klinischen Psychologie und Psychopathologie vorliegen, und in diesen Bereichen scheinen erfolgreiche Anwendungen bisher am ehesten möglich zu sein.

Zur Anwendung auf Probleme der Vorhersage ist die KFA in ihrer Grundform nur eingeschränkt tauglich. Sie erlaubt zwar die Auffindung komplexer Wechselwirkungen zwischen den Variablen. Jedoch handelt es sich dadurch, daß alle Merkmale gleichrangig behandelt werden und nicht zwischen unabhängigen und abhängigen Variablen unterschieden wird, nicht unbedingt um jene gewünschten Wechselwirkungen, die eine Vorhersage des Kriteriums

erlauben. Z.B. kann es vorkommen, daß sich zwei überfrequentierte Konfigurationen nur in ihrer Ausprägung auf dem Kriterium unterscheiden. Wenn man also nach dieser Methode eine Vorhersage versuchen wollte, würde von der gleichen Prädiktorenkonfiguration ausgehend sowohl ein hoher als auch ein niedriger Kriteriumswert vorhergesagt.

Von KRAUTH und LIENERT (1973 a, Kap. 9) wurde als Prädiktions-KFA die sogenannte Interaktionsstrukturanalyse vorgeschlagen. Erste Anwendungen des Verfahrens auf psychologische Datensätze wurden von LIENERT (1974) und in der äquivalenten Form der Mehrstichproben-KFA von ROEDER (1976) berichtet. Es handelt sich dabei um eine zweidimensionale KFA, bei der die Stufen des Kriteriums die eine, und die Konfigurationen der Prädiktoren die andere Dimension bilden. Auffällig ist bei diesem Verfahren, daß die erwarteten Häufigkeiten nicht mehr nach einem mehrdimensionalen Modell wie bei der Grundform der KFA berechnet werden, sondern daß unmittelbar von den beobachteten Häufigkeiten der Prädiktorkonfigurationen und der Kriteriumsausprägungen ausgegangen wird. Damit gewinnt das Verfahren große Ähnlichkeit mit den eingangs dieses Kapitel genannten Methoden (vor allem der ersten Variante), die ebenfalls die Prädiktorkonfigurationen direkt mit dem Kriterium in Beziehung setzen. Lediglich in der Art, in der die Beziehung zum dort intervallskalierten oder hier kategorialen Kriterium hergestellt wird, besteht ein Unterschied. M.E. besteht nur eine geringe Wahrscheinlichkeit dafür, daß sich die Interaktionsstrukturanalyse in vergleichbaren Situationen als erfolgreicher als die dort genannten Verfahren erweisen wird.

Des weiteren wird von KRAUTH und LIENERT (1973 b) vorgeschlagen, falls sich die abhängige Variable in der Prädiktions-KFA in globaler Weise vorhersagen läßt, auf Interaktionen höherer Ordnung mit Hilfe der Assoziationsstrukturanalyse, also durch systematischen, hierarchischen Einsatz der Basis-KFA (KRAUTH und LIENERT, 1973 a, Kap. 7) zu untersuchen. Dabei tritt jedoch m.E. wieder das bereits

genannte Problem auf, daß dabei keineswegs jene Wechselwirkungen entdeckt werden müssen, die für die Vorhersage des Kriteriums wichtig sind. Der gleiche Grund, der die Basis-KFA für Prädiktionsaufgaben weniger geeignet macht, entwertet also m.E. wenigstens zum Teil die im Anschluß an eine Interaktionsstrukturanalyse durchgeführten Prüfungen auf mehrdimensionale Interaktionen.

Trotz der geäußerten Bedenken und trotz der für die wenig effizienten konfigurationsanalytischen Methoden recht kleinen Stichproben soll nachfolgend eine Anwendung der KFA und der Interaktionsstrukturanalyse versucht werden. Aus technischen Gründen wurde dabei die vorgeschlagene Reihenfolge beider Verfahren umgekehrt; es werden also zuerst KFAn und anschliessend Interaktionsstrukturanalysen durchgeführt.

7.2 Ergebnisse

Es wurde das innerhalb des Systems SPSS benutzbare Programm Informationsanalyse (KÖGEL, 1977) verwendet, das neben multivariaten Informationsanalysen (vgl. ATTNEAVE, 1965, MITTENECKER und RAAB, 1973) auch KFAn ausführt. Im Rahmen der Informationsanalysen ergaben sich als zusätzliche Ergebnisse jeweils die Menge der von den Prädiktoren auf das Kriterium übertragenen Information sowie die Größe der darin enthaltenen Wechselwirkungen.

Als Stichproben wurden die Studienfächer BWL/VWL, Medizin, Math/Inf, Deutsch und Französisch untersucht; es werden nachfolgend jedoch nur die Ergebnisse der größeren Stichproben BWL/VWL, Medizin und Math/Inf berücksichtigt. Die Gesamtstichprobe von 8 Studienfächern wurde nicht analysiert, weil sich das Kriterium wegen der Notwendigkeit ganzzahliger Darstellung nicht wie bisher transformieren ließ. Einzelne Versuche zeigten, daß sich bei der Gesamtgruppe ohne Standardisierung des Kriteriums der überwiegende Anteil der Kriteriumsinformation durch das Merkmal Fachzugehörigkeit klären

ließ, während alle anderen Prädiktoren demgegenüber stark abfielen.

Es wurde - ähnlich wie in 5.2 - folgende Reihe von potentiellen Prädiktoren festgelegt:

Relevante Abiturnote (bei BWL/VWL Mathematik, bei Medizin und Math/Inf Deutsch)

Geschlecht

Soziale Schicht

Vertrauenspersonen

Depression

Verhältnis zu den Lehrern

Fachinteresse

Die Zahl der Kategorien wurde bei den Prädiktoren vorwiegend auf 3, seltener auch auf 2 oder 4 reduziert. Die Zahl der Kriteriumskategorien wurde nur bei sehr schwach besetzten Klassen reduziert; meistens ergaben sich auch hier 3 Kategorien.

Zunächst wurden KFAn mit jeweils sovielen Merkmalen - d.h. Prädiktoren und dem Kriterium - durchgeführt, wie aufgrund der Forderung nach einem fünfmal so großen Stichprobenumfang wie der Anzahl der Konfigurationen möglich war. Bei einer Zahl von jeweils 3-5 Prädiktoren, d.h. 4-6 Merkmalen, waren also Wechselwirkungen bis zu der entsprechenden Ordnung möglich. Die Analysen wurden mit sämtlichen Kombinationen der Prädiktoren durchgeführt; zusätzlich wurden weitere Analysen mit reduzierter Anzahl von Konfigurationen durch Reduzierung der Kategorien einzelner Prädiktoren und Reduzierung der Prädiktorenzahl durchgeführt. Dadurch ergaben sich bei BWL/VWL 14, bei Medizin 17 und bei Math/Inf 24 Läufe.

Es wurden jeweils der Gesamtchiquadratwert bei einer Zufallswahrscheinlichkeit von 0,05 und die einzelnen Konfigurationen nach ihren Chi-Quadrat-Komponenten asymptotisch bei einem nach der Anzahl der Vergleiche adjustierten 0,05-Niveau (vgl. KRAUTH und LIENERT, 1973 a, S. 45) statistisch beurteilt. Analog dem Vorgehen bei der Regressions- und Kontrastgruppenanalyse erfolgte keine statistische Berücksichtigung der Prädiktorenauswahl.

Da der Informationswert der Ergebnisse zu gering ist, um die Anlage von Tabellen lohnend zu machen, sollen nachfolgend die wichtigsten Ergebnisse in gering formalisierter Art aufgezählt werden.

BWL/VWL:

5 der insgesamt 14 Läufe wiesen ein signifikantes Gesamtchiquadrat auf; bei zweien der insgesamt signifikanten KFAn gab es jeweils eine signifikante Konfiguration.

Medizin:

7 der 17 Läufe hatten ein signifikantes Gesamtchiquadrat; nur in einer dieser insgesamt signifikanten KFAn traten signifikante, und zwar fünf signifikante Konfigurationen auf.

Math/Inf:

12 der 24 Läufe waren insgesamt signifikant, jedoch traten keinerlei signifikante einzelne Konfigurationen auf.

Zur Durchführung von Interaktionsstrukturanalysen war kein Rechenprogramm vorhanden, deshalb wurden die Berechnungen an einem Tischrechner unter Verwendung von Ergebnissen der vorstehend genannten KFAn durchgeführt. Wegen des damit verbundenen höheren Aufwandes wurden bei jeder Stichprobe nur jeweils die 5 Prädiktorenkombinationen berücksichtigt, die bei den KFAn die mit den geringsten Zufallswahrscheinlichkeiten verbundenen Gesamtchiquadrate aufwiesen, und die in den Informationsanalysen die höchsten Informationsmengen auf das Kriterium übertrugen. Die Signifikanzprüfungen wurden bei den Interaktionsstrukturanalysen nach dem gleichen Prinzip wie bei den KFAn durchgeführt. Nachfolgend die wichtigsten Ergebnisse:

BWL/VWL:

Alle 5 Läufe waren global signifikant; in zweien der Analysen trat jeweils eine signifikante Konfiguration auf. Es handelte sich um folgende, jeweils bessere Examenszensuren vorhersagende Konfigurationen:

a) gute Abiturnoten in Deutsch und Mathematik,

b) gute Mathematiknoten, männlich und niedrige Depression

Medizin:

Alle 5 Läufe waren global signifikant, jedoch traten keine einzelnen über- oder unterfrequentierten Konfigurationen auf.

Math/Inf:

Ebenso wie Medizin.

Insgesamt stimmten die tendenziell signifikanten Konfigurationen bei Interaktionsstrukturanalyse und KFA zum kleineren Teil überein; zum größeren Teil traten unterschiedliche Konfigurationen auf.

7.3 Diskussion

Man muß zunächst die Voraussetzungen beachten, die die vorliegenden Daten für eine Analyse durch die KFA und Interaktionsstrukturanalyse bieten: Bereits die linearen Effekte sind schwach, deutlich schwächer als bei den von LIENERT (1974) zur Demonstration benutzten Daten. Zusätzlich sind auch die Stichproben eher zu klein für eine erfolgreiche Anwendung in der vorliegenden Situation. Man mußte also annehmen, daß, wenn nicht noch neue, bisher nicht identifizierte, Effekte auftraten, mit Hilfe der hier erprobten Verfahren infolge ihrer geringen Effizienz für Daten der vorliegenden Art keine Vorhersage in wesentlichem Umfang möglich sein würde.

Die Ergebnisse bestätigen die Erwartungen und legen nahe, daß keine neuen, stärkeren Effekte in den Daten vorhanden sind. Es ist einleuchtend, daß die z.B. mit Hilfe von parametrischen, linearen Verfahren erreichten Zuordnungsquoten (vgl. 4.1) von ca. 43 % bei BWL/VWL und Medizin, sowie ca. 63 % bei Math/Inf hier keineswegs erreicht werden können.[1] Man könnte die mit Hilfe der Interaktionsstrukturanalyse zu treffende Vorhersage dadurch verbessern, daß man nicht nur die signifikanten, sondern alle Konfigurationen nutzt und für jede Prädiktorenkonfiguration die am häufigsten aufgetretene Kriteriumsklasse vorhersagt. Diese Verbesserung der Zuordnungsquote in der Analysestichprobe, die auch tatsächlich festzustellen ist, würde jedoch mit großer Wahrscheinlichkeit bei einer Kreuzvalidierung wieder verlorengehen (vgl. z.B. LUNNEBORG et al., 1967).

Wenn man die inhaltliche Bedeutung der signifikanten und tendenziell signifikanten - hier nicht mitgeteilten - Konfigurationen betrachtet, erscheint die Hypothese plausibel, daß es sich um extreme Ausprägungen von im wesentlichen linearen, kompensatorischen Effekten handelt. Dies kommt in solchen Prädiktorenkonfigurationen wie z.B. guten Deutsch- und Mathematikzensuren, oder bei LIENERT (1974) guten Ergebnissen in drei verschiedenen Tests zum Ausdruck, für die dann jeweils gute Examenszensuren bzw. Schulzensuren vorhergesagt werden. Gelegentlich werden solche Effekte dann durch weitere Merkmale moderiert, wobei jedoch meist - ähnlich wie bei der Kontrastgruppenanalyse (vgl. 6.4) - schwer zu entscheiden ist, ob solche Moderatoreffekte oder Wechselwirkungen nicht vielleicht zufällig durch hohe Kollinearitäten der Prädiktoren zustande gekommen sind. So scheinen Effekte der hier und bei LIENERT (1974) gefundenen Art relativ leicht auch mit Hilfe der Kontrastgruppenanalyse zu finden zu sein[2], und man hat dann - wenigstens bei der hier und bei

---

1) Grundsätzlich muß allerdings die Eigenschaft der KFA, meist eine größere Zahl von Individuen unklassifiziert zu lassen, nicht von besonderem Nachteil sein, wenn sich die Unklassifizierten im weniger wichtigen und interessanten Mittelbereich befinden.

2) Z.B. lassen sich die von LIENERT (1974) berichteten Ergebnisse einleuchtend auch in Form eines Baumdiagramms darstellen.

vielen Vorhersagestudien vorliegenden Art von Daten - zusätzlich den Vorteil einer größeren Effizienz.

7.4 Die Analyse von Residuen mit der KFA

Analog 6.6 wurde auch hier eine Kombination von Regressionsanalyse und KFA versucht. Mit diesem Ansatz dürfte die Frage, ob über die mit Hilfe einer Regressionsanalyse identifizierbaren linearen Effekte hinaus weitere Beziehungen der Variablen in Form nichtlinearer Effekte und komplexer Wechselwirkungen vorhanden sind, am schärfsten zu fassen sein. Hier sind - im Gegensatz zu 6.6 - auch keinerlei Einschränkungen über die Art der erfaßbaren komplexen Wechselwirkungen gültig. Positive Ergebnisse erscheinen allerdings nach den Befunden der vorigen Abschnitte nicht besonders wahrscheinlich.

Da durch die Kombination von zwei komplexen Verfahren besondere Ansprüche an den Stichprobenumfang gestellt werden, konnten nur die Studienfächer BWL/VWL und Medizin berücksichtigt werden.

Die aus den Regressionsgleichungen mit den "harten" Prädiktoren berechneten Residuen wurden so trichotomisiert, daß die Häufigkeiten in den entstandenen Klassen etwa gleich groß waren. Es ließ sich also zwischen Individuen unterscheiden, bei denen das Kriterium richtig, zu hoch oder zu niedrig vorhergesagt wurde.

Das Verfahren der KFA wurde entsprechend dem in 7.2 beschriebenen Vorgehen durchgeführt; lediglich das Merkmal "Zensur im Zwischenexamen" wurde durch die trichotomisierten Residuen ersetzt.

Bei den 14 Läufen für BWL/VWL und den 17 Läufen für Medizin ergaben sich keine Signifikanzen. Dieses Ergebnis ist eindeutig; es lassen sich keinerlei Effekte über die bereits identifizierten linearen Effekte hinaus nachweisen.

Versuchsweise wurden in den drei Analysen mit der niedrigsten Zufallswahrscheinlichkeit die Konfigurationen mit den ebenfalls niedrigsten Zufallswahrscheinlichkeiten aufgesucht und interpretiert. Die Überlegungen dabei ließen den Eindruck entstehen, daß sich die KFA oder vielleicht besser die Interaktionsstrukturanalyse bei entsprechender Datenstruktur mit Gewinn zur Analyse von Residuen oder von Differenzwerten nach GHISELLI (1960, vgl. auch 5.4) einsetzen ließe.

Zusammenfassend läßt sich für Kap. 7 feststellen, daß sich mit Hilfe von konfigurationsanalytischen Methoden nach KRAUTH und LIENERT (1973 a) keine wesentlichen, neuen, bisher nicht identifizierbaren Effekte sichern ließen. Erwartungsgemäß erwiesen sich die benutzten konfigurationsanalytischen Methoden als weniger effizient als die bisher verwendeten Verfahren, und aus weiteren grundsätzlichen Erwägungen heraus erschien es unwahrscheinlich, daß konfigurationsanalytische Methoden für Daten der vorliegenden Art als gleichwertige Alternativen zu Regressionsanalyse und Kontrastgruppenanalyse betrachtet werden können. Am aussichtsreichsten erscheint die Anwendung von konfigurationsanalytischen Methoden auf von den linearen Effekten der Prädiktoren bereinigte Residuen.

8. Schlußfolgerungen

In der vorliegenden Untersuchung, die allgemein gesehen mit Methodenproblemen innerhalb der Hochschulforschung befaßt ist, wurde nach Möglichkeiten gesucht, die Vorhersage des Studienerfolgs zu verbessern. Den Ausgangspunkt bildete die Überlegung und Vermutung, daß die - wie allgemein bekannt - nur geringe Vorhersagbarkeit des Studienerfolgs durch die Beschränkung der bisherigen Forschung auf lineare Vorhersagemodelle verursacht sein könnte. Es sind nämlich komplexe Zusammenhänge zwischen Merkmalen - wie z.B. paradigmatisch von MEEHL (1950) beschrieben - möglich, die bei den im wesentlichen auf paarweiser Analyse der Variablen basierenden linearen Modellen nicht berücksichtigt werden und deshalb bisher nicht für die Vorhersage des Studienerfolgs genutzt wurden. Die Hauptfrage dieser Studie lautet also: Ist eine bessere Vorhersage des Studienerfolgs als bisher möglich, wenn man Methoden zur Analyse der Daten benutzt, die auch solche komplexen Interaktionen zwischen den Variablen zur Entwicklung von Vorhersagestrategien nutzen.

Verschiedene Methoden der Datenanalyse - rangierend vom herkömmlich üblichen linearen bis zum konfigurationsanalytischen Modell - wurden in exemplarischer Anwendung auf einen vorliegenden Datensatz einem kritischen Vergleich unterzogen. Dabei wurde die Eignung der Methoden zur Transformation jeglicher Strukturen der Daten in unmittelbar verständliche und für die Lösung von Problemen nützliche Form evaluiert.

Insgesamt gesehen konnte eine relativ hochgespannte Erwartung, daß sich mit Hilfe von komplexen Interaktionen zwischen den Variablen die Vorhersage des Studienerfolgs unmittelbar auf ein wesentlich höheres quantitatives Niveau heben ließe, nicht bestätigt werden.

Darüber hinaus ist jedoch eine Reihe von Befunden zu nennen, die sich bei der vorgenommenen exemplarischen Demonstration der Einsatzmöglichkeiten verschiedener Analysemethoden als wesentlich für die Methodologie des Problemgebietes ergaben:

Zu Beginn wurde aufgrund verschiedener Überlegungen festgestellt, daß, obwohl ein niedrigeres Meßniveau als Intervallskalenniveau vorlag, die Skalenqualität der Merkmale kein grundsätzliches Hindernis für die Anwendung parametrischer Methoden darstellt. Allerdings ist mit Auswirkungen durch die Unvollkommenheiten der Meßverfahren auf die Ergebnisse zu rechnen, und diese Unvollkommenheiten dürften auch zum Teil für die geringen Vorhersagevaliditäten der Prädiktoren verantwortlich sein.

Bei der Durchführung von multiplen Regressionsanalysen wurde von einer Vorhersage durch die Abiturnoten allein ausgegangen; die weiteren Prädiktoren wurden den Abiturnoten in zwei Schritten hinzugefügt. Es mußten dabei die Studienfächer einzeln analysiert werden; die Zusammenfassung aller Studienfächer erwies sich trotz fachweiser Standardisierung aller Merkmale außer zur Demonstration fachübergreifender Tendenzen als nicht möglich. Bei allen Studienfächern zeigten sich die Abiturnoten als die stärksten Prädiktoren; die den Abiturnoten hinzugefügten Variablen brachten in den meisten Fächern eine Verbesserung der Vorhersage, und ihre Effekte ließen sich vielfach in Übereinstimmung mit bekannten Eigenschaften der betreffenden Studienfächer interpretieren. Eine Verwendung dieser den Abiturnoten hinzugefügten Variablen für praktische Auslesezwecke könnte allerdings wohl nur in sehr wenigen Fällen gerechtfertigt werden; die Ergebnisse sind also mehr als grundsätzlich denn als anwendungsorientiert zu verstehen.

Der Einsatz von Diskriminanzanalysen als Alternativen zu multiplen Regressionsanalysen zeigte, daß die größeren Freiheitsgrade des diskriminanzanalytischen Modells zur Anpassung an die Daten nicht in wesentlichem Umfang genutzt wurde. Die Auffassung des Kriteriums innerhalb der Regressionsanalysen als eindimensionale Intervallskala ist also gerechtfertigt, und die multiple Regressionsanalyse als das in dieser Situation effizientere Modell verdient den Vorzug gegenüber der Diskriminazanalyse.

Die Entdekcung von Ausleseeffekten der Studienfächer auf die Abiturzensuren führte zu der Unterscheidung von zwei verschiedenen für Vorhersagezwecke nutzbaren Komponenten innerhalb der Abitur-

noten: Die erste Komponente sagt die Wahl des Studienfachs, die zweite Komponente den Examenserfolg in einem bestimmten Fach voraus. Dabei umfaßt die erste Komponente vorwiegend jene Schulfächer, die man als die dem Studienfach entsprechenden Schulfächer für die Vorhersage gerade des Erfolgs in dem betreffenden Studienfach als wesentlich ansehen müßte. Dahingegen erfolgt die tatsächliche Vorhersage des Examenserfolgs in Form der zweiten Komponente vor allem durch solche - jedenfalls nach allgemein vorherrschender Ansicht - Abiturfächer, die eigentlich weniger relevant zur Vorhersage des Examenserfolgs sind. Offensichtlich ist die Varianz der "eigentlich" relevanten Abiturnoten nach der Wahl des Studienfachs im wesentlichen "verbraucht", so daß für die empirisch zu sichernde Vorhersagevalidität der Abiturnoten für den Studienerfolg nur noch die restlichen, "eigentlich" weniger relevanten Abiturfächer übrig bleiben. In der Praxis muß die Validität der Abiturnoten also in zwei Stufen gesehen werden: Zunächst in bezug auf die Wahl des Studienfachs und sodann in bezug auf den Erfolg innerhalb des gewählten Fachs. Es liegt die Vermutung nahe, daß die theoretisch "wahre" Validität der Abiturnoten für die Vorhersage des Examenserfolgs - falls keine Auslese durch die Wahl des Faches stattfände - wesentlich höher und bei anderen Abiturfächern läge als die empirisch gefundene Validität. Dieser Punkt ist sicherlich für die Diskussion von Auslesestrategien von Bedeutung, da durch eine von außen gesetzte Auslese die Selbstauslesestrategien der Abiturienten verändert und wenigstens zum Teil außer Kraft gesetzt würden.

Es wurde ein Modell vorgeschlagen, das nicht nur die Zensuren der Studenten, die sich einem Examen unterzogen, als Kriterium einbezieht, sondern auch die Tatsache eines Studienabbruchs berücksichtigt. Dabei sind zwei Typen von Ergebnissen möglich, die von unterschiedlicher Auswirkung auf die Bewertung der bisherigen, nur Teilaspekte des Examenserfolgs einbeziehenden, Ergebnisse sind.

Verschiedene Prüfungen daraufhin, ob generell ein lineares, und speziell das regressionsanalytische Modell zur Analyse der vorliegenden Daten geeignet ist, bestätigten im ganzen die Anwendbarkeit des regressionsanalytischen Modells. Systematische Abweichungen

von den Modellannahmen zeigten sich lediglich in Heteroskedastizitäten der Residuen, die jedoch zu gering im Ausmaß waren, um das angewendete Modell zu entwerten oder den Einsatz des "generalized least squares estimator" (AITKIN, nach JOHNSTON, 1963), der solche Heteroskedastizitäten zu berücksichtigen gestattet, nahezulegen.

Von zwei hier erprobten Techniken der Moderatoranalyse erbrachte die Methode von SAUNDERS (1956) keinerlei und die Methode von GHISELLI (1960) schwache, aber systematisch auftretende Effekte. Die hier anscheinend grundsätzlich bestehende Geltung des Modells von GHISELLI läßt die systematische Entwicklung solcher Moderatoren aussichtsreich erscheinen (vgl. GHISELLI, 1972).

In den Regressionsgleichungen waren einzelne Suppressorvariablen aufgetaucht, von denen in der Vergangenheit grundsätzlich ähnlich wie von Moderatoren eine Überwindung der "prediction ceiling" erhofft wurde[1)], und ihre Effekte ließen sich in einleuchtender Weise interpretieren. Jedoch erschien die heute allgemein vorherrschende Meinung bestätigt, daß sich Suppressoren schwerlich in systematischer Weise suchen und einsetzen lassen.

Der Einsatz der Kontrastgruppenanalyse (auch AID genannt, vgl. z.B. SONQUIST et al., 1974) und der Konfigurationsfrequenzanalyse bzw. Interaktionsstrukturanalyse (vgl. z.B. KRAUTH und LIENERT, 1973 a) sollte die Daten auf das Vorhandensein von mehrdimensionalen Interaktionen - wie z.B. von MEEHL (1950) beschrieben - prüfen. Die Ergebnisse waren im ganzen - wie bereits eingangs erwähnt - eher negativ. AID erwies sich zwar als in etwa gleich effizient wie die multiple Regressionsanalyse; jedoch wurde die Interpretation von Wechselwirkungen durch die Kollinearitäten der Prädiktoren verhindert, bzw. waren die Interaktionen zu schwach, um sich in überzeugender Weise darzustellen. KFA und Interaktionsstrukturanalyse - als deutlich geringer effiziente Verfahren im Vergleich zur multiplen Regressionsanalyse - konnten schließlich kaum signifikante Zusammenhänge zwischen den Variablen und damit auch keine komplexen Interaktionen aufdecken.

---

1) Beide Arten von Variablen - Suppressoren wie auch Moderatoren - wurden z.B. gelegentlich als "Irrlichter" der Vorhersageforschung bezeichnet.

Wenn sich in dieser Studie lineare Effekte als die entscheidenden Strukturen innerhalb der analysierten Daten erwiesen und keine komplexen, konfigurationalen Zusammenhänge von überzeugender Stärke beobachtet werden konnten, ist dies zunächst eine Bestätigung der Angemessenheit der bisher ausschließlich verwendeten linearen Modelle. Weiterhin muß die hier festgestellte Abwesenheit von komplexen Interaktionen keineswegs eine endgültige Feststellung sein:

1. In anderen als den hier analysierten Datensätzen könnten, auch wenn die Wahrscheinlichkeit dafür vielleicht gering ist, doch einmal Effekte auftauchen, die sich nicht mit linearen Methoden erfassen lassen.
2. Auch innerhalb der hier verwendeten Daten wurde in Tendenzen eine mögliche Anwendbarkeit des Moderatormodells von GHISELLI faßbar. Eine systematische Suche und Entwicklung solcher Moderatoren könnte vielleicht Tendenzen in anwendbare Effekte verstärken.

Insgesamt dürfte in der Suche nach nichtlinearen Effekten nach wie vor die hauptsächliche Hoffnung liegen, von der Seite der Methodologie her zu einer Verbesserung der Vorhersagetechniken bzw. zu einer Überwindung des "prediction ceiling" beizutragen. So sollte m.E. vor allem folgenden Ansätzen als Ergänzungen der linearen Modelle auch in der Zukunft besondere Aufmerksamkeit zugewendet werden:

- Entwicklung und systematische Suche von Moderatoren nach der Methode von GHISELLI (1972)
- Anwendung der Kontrastgruppenanalyse in einer Form, die neu entwickelte Signifikanztests bereits während der Durchführung der Analyse - also innerhalb des Rechenprogramms - berücksichtigt
- Analyse von Residuen oder Differenzwerten nach GHISELLI, die aus Regressionsgleichungen gewonnen wurden, mit Hilfe der KFA oder Kontrastgruppenanalyse; d.h. Kombination des linearen mit einem nichtlinearen Modell.

LITERATURVERZEICHNIS

ABRAHAMS, Norman M.
Off-quadrant comment, J. appl. Ps., 1969, 53, 66-68

ABRAHAMS, N.M. & ALF, Jr.
Pratfalls in moderator research. J. appl. Ps., 1972 a, 56, 245-251

ABRAHAMS, N.M. & ALF, Jr.
Reply to Dunnette's "Comments on Abrahams and Alf's "Pratfalls in moderator research". J. appl. Ps., 1972 b, 56, 257-261

ADAMS, E.W., R.F. FAGOT & R.E. ROBINSON
A theory of appropriate statistics, Psychometrics, 1965, 30, 99-127

AHRENS, Heinz
Varianzanalyse, Berlin 1968

ALLERBECK, Klaus
Datenverarbeitung in der empirischen Sozialforschung, Stuttgart, 1972

AMELANG, M., SOMMER, E. & BARTUSSEK, D.
Persönlichkeitsstruktur und Studienrichtung, Psychologische Beiträge, 1971, 13, 7-25

AMELANG, M. & T. HOPPENSACK
Persönlichkeitsstruktur und Hochschulbesuch, Ps. Erz. Unt., 1977 a, 19, 193-204

AMELANG, M. & T. HOPPENSACK
Persönlichkeitsstruktur und Hochschulbesuch- Merkmalsveränderungen während des Hochschulbesuchs, Psychol. Beiträge, 1977 b, 19, 161-188

ANSCOMBE, F.J. & J.W. TUKEY
The examination and analysis of residuals, Technometrics, 1963, 5, 141-160

APENBURG, E., GROSSKOPF, R. & SCHLATTMANN, H.
Orientierungsprobleme und Erfolgsbeeinträchtigung bei Studierenden, Bericht über eine Befragung an 4500 Studenten. Universität des Saarlandes, 1974

ARMSTRONG, J.S. & J.G. ANDRESS
Exploratory analysis of marketing data: Trees vs. regression, J. marketing research, 1970, 7, 487-492

ATTNEAVE, F.
Informationstheorie in der Psychologie, Bern, 1965

BAKER, B.O., C.D. HARDYCK & L.F. PETRINOVICH
Weak measurements vs. strong statistics, Ed. ps. meas., 1966, 26, 291-309

BARTENWERFER, H. & GIESEN, H.
Pilotstudie über die Beobachtung und Analyse von Bildungslebensläufen, Frankfurt/Main: Deutsches Institut für Internationale Pädagogische Forschung, 1973.

BARTUSSEK, Dieter
Eine Methode zur Bestimmung von Moderatoreffekten, Diagnostika, 1970, 16, 57-76

BAUER, Adam
Ein Verfahren zur Messung des für Bildungsverhalten relevanten sozialen Status, Frankfurt/M: Deutsches Institut für Internationale Pädagogische Forschung, 1972

BECKMANN, D., MOELLER, M.L., RICHTER, H. & SCHEER, J.W.
Studenten: Urteile über sich selbst, über ihre Arbeit und über die Universität, Aspekte Verlag, Frankfurt/Main, 1972.

BIERVERT, B., M. DIERKES & A. WALZEL
Automatischer Split zur optimalen Kombination erklärender Faktoren, Kölner Zt. Soziol. u. Sozialps. 1969, 21, 304-323

BINDER-KRAUTHOFF, Kristine
Verhaltensrelevante Aspekte von Abiturnoten bei Studienwahl und Studienerfolg, Sociologia Internationalis, 1971, 9, 71-85.

BOCHNIK, DONIKE & PITTRICH
Numerus clausus in der Medizin, 1974.

BRADLEY, J.V.
Distribution-free statistical tests, Englewood Sliffs/ N.J., 1968

CATELL, R.B.
The scree test for the number of factors, Multivar. behav. res., 1966, 1, 245-276

CLAUDY, John, G.
A comparison of 5 variable weighting procedures. Educational and Psychological Measurement, 1972, 32, 311 - 322.

COHEN, Jacob
Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 70, 426-443.

CONGER, Anthony, J.
An analysis of Ghiselli's moderator variable, Ps. reports, 1969, 25, 519-527

CONGER, A.J. & D.N. JACKSON
Suppressor variables, prediction and the interpretation of psychological relationships, Ed. ps. meas., 1972, 32, 579-599

CONGER, A.J. & D.N. JACKSON
A revised definition for suppressor variables, Ed. ps. meas., 1974, 34, 35-46

COOLEY, W.W. & P.R. LOHNES
Multivariate data analysis, New York, 1971

DARLINGTON, Richard B.
Multiple regression in psychological research and practice, Ps. Bull., 1968, 69, 161-182

DIGMAN, John M.
Interaction and non-linearity in multivariate experiment, in: CATELL, R.B. (Hg) Handbook of multivariate experimental psychology, Chicago, 1966

DRAPER, N.R., I. GUTTMAN & H. KANEMASU
The distribution of certain regression statistics, Biometrica, 1971, 58, 1-4

DRAPER, N.R. & H. SMITH
Applied regression analysis, New York, 1966

DUNNETTE, Marvin D.
Comments on Abrahams and Alf's "Pratfalls in moderator research" J. appl. ps., 1972, 56, 252-256

EHLERS, Th. & GIESEN, H.
Eine Längsschnittuntersuchung zur Vorhersage des Studienerfolgs bei Marburger Psychologiestudenten. Ber. aus dem Inst. f. Psychol. d. Phil.-Univ. in Marburg/Lahn, 1971, Nr. 27.

FARRAR, D.E. & R. GLAUBER
Multicollinearity in regression analysis: The problem re-visited, Rev. of economics and statistics, 1967, 49, 92-107

FESTERLING, K.-H.
Akademische Abschlußprüfungen. Baden-Württemberg in Wort und Zahl. Statist. Monatshefte, 19, 1971, 337-342.

FISCH, R., SCHMALT, H.D. & FISCH, H.
Ein Meßverfahren zur Bestimmung der Intensität und Extensität der Leistungsmotivation. Unveröff. Bericht, Psychol. Inst. der Univ. des Saarlandes, 1970

FOREHAND, G.A. & MCQUITTY, L.L.
Configurations of factor standings as predictors of educational achievement, Ed. ps. meas., 1959, 19, 31-43

GAENSSLEN, H. & W. SCHUBÖ
Einfache und komplexe statistische Analyse, München, 1973

GHISELLI, Edwin E.
Differentiation of individuals in terms of their predictability, J. appl. Ps., 1956, 40, 374-377

GHISELLI, Edwin E.
Differentiation of tests in terms of the accuracy with which they predict for a given individual, Ed. ps. meas., 1960, 20, 615-684

GHISELLI, Edwin E.
Moderating effects and differential reliability and validity, J. appl. ps., 1963, 47, 81-86

GHISELLI, Edwin E.
The validity of occupational aptitude tests, New York, 1966

GHISELLI, Edwin E.
Moderating heteroscedasticity, Ed. ps. meas., 1967, 27, 581-590

GHISELLI, Edwin E.
Comment on the use of moderator variables, J. appl. ps., 1972, 56, 270

GROSS, A.L., J. FAGGEN-STECKLER & K. MCCARTHY
Statistical procedures for evaluating the practical utility of a moderator approach to prediction, J. app. ps., 1974 a, 59, 578-582

GROSS, A.L., J. FAGGEN & K. MCCARTHY
The differential predictability of the college performance of males and females, Ed. ps. meas., 1974 b, 34, 363-365

GROSS, A.L. & Wen-huey SU
Defining a "fair" or "unbiased" selection model, J. appl. ps., 1975, 60, 345-351

GUILFORD, J.P.
Fundamental statistics in psychology and education, New York, 1965

HAYS, W.L.
Statistics for psychologists, New York, 1963

HOBERT, R. & DUNETTE, M.D.
Development of moderator variables to enhance the prediction of managerial effectiveness, J. appl. ps., 1967, 51, 50-64

HOPE, Keith
Methoden multivariater Analyse, Weinheim, 1975

HORST, Paul
The prediction of personal adjustment, Social science research council bulletin, Nr. 48, New York, 1941

HORST, Paul
Pattern analysis and configural scoring, J. clin. ps., 1954, 10, 3-11

ICPR
OSIRIS II - OS users manual, Ann Arbor, Mich., 1971

INGENKAMP, K. (Hg)
Die Fragwürdigkeit der Zensurengebung, Weinheim, 1976

JACQUEZ, J.A., F.J. MATHER & C.R. CRAWFORD
Linear regression with nonconstant, unknown error variances, Biometrics, 1968, 24, 607-626

JOHNSTON, J.
Econometric methods, New York, 1963

JONES, Marshall B.
Moderated regression and equal opportunity, Ed. ps. meas., 1973, 33, 591-602

KASS, G.V.
Significance testing in Automatic Interaction Detection (AID), Appl. statist., 1975, 24, 178-189

KATH, G., OEHLER, C. & REICHWEIN, R.
Studienweg und Studienerfolg, Institut für Bildungsforschung in der Max-Planck-Gesellschaft zur Förderung der Wissenschaften, Berlin, 1966.

KERLINGER, F.N. & E.J. PEDHAZUR
Multiple regression in behavioral research, New York, 1973

KESELMAN, H.J.
Multiple testing and type I errors, Am. ps., 1974, 29, 778-779

KLEIN, S.P., D.A. ROCK & F.R. EVANS
The use of multiple moderators in academic prediction, J. ed. meas., 1968, 5, 151-160

KÖGEL, Dieter
Benutzerhandbuch Informationanalyse, Rechenzentrum der Universität zu Köln, 1977

KOPFMANN, Gerhard
Zusammenhang zwischen Abiturnoten von Physikstudenten und ihrer Leistung an der TH München, Zeitschrift für experimentelle und angewandte Psychologie, 1969, 16, 434-451

KRAUTH, J. & G.A. LIENERT
Die Konfigurationsfrequenzanalyse, Freiburg, 1973 a

KRAUTH, J. & G.A. LIENERT
Zum Nachweis syndromgenerierender Symptominteraktionen in mehrdimensionalen Kontingenztafeln (Interaktionsstrukturanalyse), Biometrische Zeitschr., 1973 b, 15

LANCASTER, H.O.
The chi- squared distribution, New York, 1969

LEE, M.C.
Interactions, configurations and non-additive models, Ed. ps. meas., 1961, 21, 797-805

LIENERT, G.A.
Testaufbau und Testanalyse, Weinheim, 1967

LIENERT, G.A.
Die "Konfigurationsfrequenzanalyse" als Klassifikationsmittel in der klinischen Psychologie. In: IRLE (Hg.) Bericht über den 26. Kongreß der deutschen Gesellschaft für Psychologie, 1968, Göttingen, 1969, 244-253

LIENERT, G.A.
Die Konfigurationsfrequenzanalyse I, Zeitschr. klin. Ps. Psychoth., 1971, 19, 99-115

LIENERT, G.A.
Verteilungsfreie Methoden in der Biostatistik, Bd. 1, Meisenheim 1973

LIENERT, G.A.
Die Konfigurationsfrequenzanalyse als Prädiktionsmodell in der angewandten Psychologie, in; ECKENSBERGER et al. Bericht über den 28. Kongreß der deutschen Gesellschaft für Psychologie in Saarbrücken 1972, Göttingen, 1974

LIENERT, G.A. & J. KRAUTH
Configural frequency analysis as a statistical tool foor defining types, Ed. ps. meas., 1975, 35, 231-238

LÖSEL, F. & W. WÜSTENDÖRFER
Zum Problem unvollständiger Datenmatrizen in der empirischen Sozialforschung, Kölner Z. Soziol. Sozialps., 1974, 26, 536-552

LORD, F.M.
On the statistical treatment of football numbers, Am. ps., 1953, 8, 750-751

LORD, F.M.
Efficiency of prediction when a regression equation from one sample is used in a new sample, Princeton, N.J., Educational Testing Service, 1950, nach: BROGDEN, H.E. Statistical theory and research design, Ann. rev. of Ps., 1954, 5, 381

LUBIN, A.
Statistics, in: Annual review of psych., Palo Alto/Calif., 1962

LUBIN, A. & H.G. OSBURN
A theory of pattern analysis for the prediction of a quantitative criterion, Psychometrika, 1957, 22, 63-73

LUNNEBORG, C.E. & LUNNEBORG, P.W.
Pattern prediction of academic success, Ed. ps. meas., 1967, 27, 945-952

MARKS, M.R.
Two kinds of regression weights that are better than betas in crossed samples. Paper read to APA Convention, September 1966.

McNEMAR, Q.
Psychological statistics, New York, 1962

McNEMAR, Q.
Moderation of a moderator technique, J. appl. ps., 1969, 53, 69-72

MEEHL, Paul E.
Configural scoring, J. cons. ps., 1950, 14, 165-171

MEREDITH, W.M., C.H. FREDERIKSEN & D.H. MCLAUGHLIN
Statistics and data analysis, Ann. rev. of Ps., 1974, 25, 453-505

MITTENECKER, E. & E. RAAB
Informationstheorie in der Psychologie, Göttingen, 1973

MORGAN, J.N. & J.A. SONQUIST
Problems in the analysis of survey data, and a proposal, J. of the am. stat. ass., 1963, 58, 415-434

NICHOLSON, G.E.Jr.
Prediction in future samples, in: WLKIN, I. et al. (Hg.), Contributions to probability and statistics, Stanford, University Press, 1960

NIE, N. et al.
SPSS - Statistical Package for the Social Sciences, New York, 1970

ORLIK, Peter
Kritische Untersuchungen zur Begabtenförderung, Meisenheim am Glan: Hain, 1967

PINDER, Craig C.
Statistical accuracy and practical utility in the use of moderator variables, J. appl. ps. 1973, 57, 214-221

PRESS, S. James
Applied multivariate analysis, New York, 1972

ROCK, A.D., J.L. BARONE & R.L. LINN
A fortran computer program for a moderated stepwise prediction system, Ed. ps. meas., 1967, 27, 709-713

ROCK, A.D. & F.R. EVANS
Predicting multiple criteria of creative achievements with moderator variables, J. of ed. meas., 1969, 6, 229-235

ROEDER, B.
Parameterfreier Vergleich verschiedener Konfigurationsfrequenzanalysen, Kölner Z. Soziol. Sozialps., 1976, 28, 152-155

ROMINE, H.B., J.A. DAVIS & W.S. GEHMAN
The interaction of learning, personality, traits, ability, and environment, Ed. ps. meas., 1970, 30, 337-347

RUNDQUIST, Edward A.
The prediction ceiling, Personnel Ps., 1969, 22, 109-116

SAUNDERS, D.R.
Moderator variables in prediction, Ed. Ps. meas., 1956, 16, 209-222

SAVAGE, I.R.
Non-parametric statistics, J. am. stat. assoc., 1957, 52, 331-344

SCHLATTMANN, H. & E. APENBURG
Psychosoziale Bedingungen von Studienleistungen, Saarbrücker Studien zur Hochschulentwicklung, Heft 27, Saarbrücken, 1977

SCHMIED, Dieter
Möglichkeiten und Probleme der Studienerfolgsprognose, Diss. an der TU Berlin, 1971.

SCHÜTZ, Erich
Studien- und Schulleistungen, Deutsches Ärzteblatt, 1968, 9, 510-516

SCOTT, A.J. & M. KNOTT
An approximate test for use with AID, Appl. stat., 1976, 25, 103-106

SELVIN, Hanan C.
In BJERSTEDT, R., A design for sociology: Scope, objectives, and methods, Philadelphia, 1969

SIEGEL, S.
Nonparametric statistics, New York, 1956

SIGNORELLI, Anthony
Statistics- tool or master of the psychologist? Am. ps., 1974, 29, 774-777

SONQUIST, J. A. & J. N. MORGAN
The detection of interaction effects, Ann Arbor/Mich., 1964

SONQUIST, J., E. BAKER & J.N. MORGAN
Searching for structure, Ann Arbor, Institute for Social Research, 1973

STEIN, C.
Multiple regression, in: OLKIN, I. et al. (Hg.), Contributions to probability and statistics, Stanford, Stanford university press, 1960

STEVENS, S.S.
Mathematics, measurement and psychophysics, in: STEVENS (Hg.), Handbook of experimental psychology, New York, 1951

STEVENS, S.S.
Measurement, psychophysics and utility, in: CHURCHMAN et al. (Hg.), Measurement: definitions and theories, New York, 1959

TENT, L., W. FINGERHUT & H. LANGFELDT
Quellen des Lehrerurteils, Weinheim, 1976

TROST, Günter
Vorhersage des Studienerfolgs, Braunschweig, 1975

TUKEY, John
The future of data analysis, Annals of mathematical statistics, 1962, 33, 1-67

UHL, N. & T. EISENBERG
Predicting shrinkage in the multiple correlation coefficient, Ed. ps. meas., 1970, 30, 487-489

VAN DE GEER, John P.
Introduction to multivariate analysis for the social sciences, S. Francisco, 1971

VELICER, Wayne F.
The moderator variable viewed as heterogeneous regression, J. appl. ps., 1972 a, 56, 266-269

VELICER, Wayne. F.
Comment on the general inapplicability of Ghiselli's moderator system for two predictors, J. appl. ps., 1972 b, 56, 262-265

ZEDECK, Sheldon
Problems with the use of "moderator" variables, Psych. bull, 1971 a, 76, 295-310

ZEDECK, Sheldon
Identification of moderator variables by discriminant analysis in a multipredictable group validation model, J. appl. ps., 1971 b, 55, 364-371

ZEDECK, Sheldon
Comparison of "joint moderators" in three prediction techniques, J. appl. ps., 1971 c, 55, 234-240

LEBENSLAUF

30.1.1942	geboren in Bielefeld, Westfalen
1948 - 1952	Besuch der Volksschule in Rödinghausen, Kr. Herford
1952 - 1955	Besuch des Gymnasiums Bünde, Kr. Herford
1955 - 1961	Besuch des Helmholtz-Gymnasiums, Bielefeld Ablegung der Reifeprüfung 1961
1961 - 1966	Musikstudium an der Musikhochschule und der Freien Universität Berlin
1966 - 1969	Psychologiestudium an der Universität Hamburg, Ablegung des Diplom-Vorexamens 1969
1969 - 1971	Psychologiestudium an der Universität Düsseldorf, Ablegung der Diplom-Hauptprüfung 1971
Juni - Dezember 1971	Wissenschaftliche Hilfskraft am Psychologischen Institut der Universität Düsseldorf
Jan. 1971 - Dez.1972	Wissenschaftlicher Mitarbeiter am Deutschen Institut für Internationale Pädagogische Forschung, Frankfurt, Abteilung für Statistik und Methodenforschung
Jan. 1973 - Jan.1975	Wissenschaftlicher Mitarbeiter der Universität des Saarlandes im Forschungsprojekt "Orientierungsprobleme und Erfolgsbeeinträchtigung bei Studierenden"
seit Febr. 1975	Verwalter der Stelle eines wissenschaftlichen Assistenten an der PH Rheinland, Abteilung Aachen, im Fach Psychologie

Fotodruck Mainz, Aachen, Malteserstraße 9, am Ponttor, Telefon (02 41) 2 73 05